

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2002, 15:28:23 ; Search time 50.82 Seconds  
(without alignments)  
6481.592 Million cell updates/sec

Title: US-09-910-186A-7  
Perfect score: 1341  
Sequence: 1 gaattcacgatggccaacaa.....ggaccgaatagtaagaattc 1341

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues  
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	263	19.6	1330	1 US-08-480-604A-22	Sequence 22, Appl
2	263	19.6	1330	2 US-08-405-496A-22	Sequence 22, Appl
3	263	19.6	1330	4 US-08-915-136-22	Sequence 22, Appl
4	263	19.6	1402	1 US-08-480-604A-25	Sequence 25, Appl
5	263	19.6	1402	2 US-08-405-496A-25	Sequence 25, Appl
6	263	19.6	1402	4 US-08-915-136-25	Sequence 25, Appl
7	131.4	9.8	1359	1 US-07-618-312A-3	Sequence 3, Appl
8	131.4	9.8	1359	1 US-08-280-228-3	Sequence 3, Appl
9	123	9.2	3891	1 US-08-480-604A-27	Sequence 27, Appl
10	123	9.2	3891	2 US-08-405-496A-27	Sequence 27, Appl
11	123	9.2	3891	4 US-08-915-136-27	Sequence 27, Appl
12	70.6	5.3	5163	3 US-08-700-651-1	Sequence 1, Appl
13	70.6	5.3	5163	3 US-08-928-361B-4	Sequence 4, Appl
14	70.6	5.3	5318	3 US-08-700-651-2	Sequence 2, Appl
15	70.6	5.3	5318	3 US-08-928-361B-2	Sequence 3, Appl
16	64	4.8	5511	3 US-08-928-361B-2	Sequence 2, Appl
17	64	4.8	7334	3 US-08-928-361B-1	Sequence 1, Appl
18	55.4	4.1	1359	1 US-07-618-312A-1	Sequence 1, Appl
19	55.4	4.1	1359	1 US-08-110-786A-7	Sequence 7, Appl
20	55.4	4.1	1359	1 US-08-280-228-1	Sequence 1, Appl
21	55.4	4.1	1858	1 US-08-668-381A-6	Sequence 6, Appl
22	43.4	3.2	1430	1 US-08-276-452A-25	Sequence 25, Appl
23	43.4	3.2	1430	2 US-08-798-744-25	Sequence 25, Appl
24	40.4	3.0	1690	1 US-08-276-452A-24	Sequence 24, Appl
25	40.4	3.0	1690	2 US-08-798-744-24	Sequence 24, Appl
26	38.4	2.9	198	5 PCT-US95-10668-3	Sequence 3, Appl
27	38.4	2.9	198	5 PCT-US95-10668-4	Sequence 4, Appl

28	36.8	2.7	9046	1 US-08-227-536-1	Sequence 1, Appl
29	36.8	2.7	9046	5 PCT-US95-04682-1	Sequence 1, Appl
30	36.6	2.7	6826	3 US-09-024-020B-8	Sequence 8, Appl
31	36.6	2.7	6826	4 US-09-425-043-8	Sequence 8, Appl
32	36	2.7	1149	4 US-09-422-487-6	Sequence 6, Appl
33	36	2.7	18596	4 US-09-318-448-11	Sequence 11, Appl
34	35.6	2.7	3378	1 US-07-972-791-1	Sequence 1, Appl
35	35.2	2.6	198	5 PCT-US95-10668-1	Sequence 1, Appl
36	35.2	2.6	198	5 PCT-US95-10668-2	Sequence 2, Appl
37	35	2.6	4291	4 US-09-351-200-1	Sequence 1, Appl
38	35	2.6	5977	3 US-09-024-020B-1	Sequence 1, Appl
39	35	2.6	5977	4 US-09-425-043-1	Sequence 1, Appl
40	35	2.6	6007	3 US-09-024-020B-2	Sequence 2, Appl
41	35	2.6	6007	4 US-09-425-043-2	Sequence 2, Appl
42	35	2.6	6556	3 US-09-024-020B-7	Sequence 7, Appl
43	35	2.6	6556	4 US-09-425-043-7	Sequence 7, Appl
44	35	2.6	6586	3 US-09-024-020B-43	Sequence 43, Appl
45	35	2.6	6586	4 US-09-425-043-43	Sequence 43, Appl

ALIGNMENTS

RESULT 1  
US-08-480-604A-22  
; Sequence 22, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027









```

; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIORITY DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1402 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1386
; US-08-480-604A-25

```

Query Match 19.6%; Score 263; DB 1; Length 1402;  
Best Local Similarity 53.9%; Pred. No. 1.7e-66;  
Matches 639; Conservative 0; Mismatches 525; Indels 21; Gaps 4;

QY	15	caacaataacaattccggaattccggaataatcatcctggaacctgcgtttacaagaaca	74
Db	96	CACtGAATACATCAAGAAGACATCATCAATACCTCCATCTTGAACCTGCGGTACGAATCCAA	155
QY	75	caactcgatcgatctctcgtgttacggtgctaaagtgaagtatacgacgcggtgtc- ----	129
Db	156	TCACCTGATCGACCTCTCTCGCTACGCTTCCAAATCAACATCGGTTCTAAAGTTAACTT	215
QY	130	-gaactgaatgacaagaacacaggttcaaacctgacctctcttcgcgttaacttaagatccggtg	188
Db	216	CGATCGGATCGACAGAAGATCAGATCCAGCTGTTCATCTGGAATCTTCCAAAATCGAAGT	275
QY	189	tactcgaatcgagaacatcatcttcaactccgtattctcctggactctctgttctctctcg	248
Db	276	TATCTTGAGAGATGCTATCGTATACAACTCTATGTACGAAAATCTCTCCACCTCTCTCTG	335
QY	249	gattcgtattccggaatacaagaacgcggtatccagaattacatcccaacaatgaatacac	308
Db	336	GATCCGTATCCGGAATACTTCAACTCC- ----ATCTCTCTGAACAATGAATACAC	386
QY	309	catcatcaactgcataagaataactctcgtgttggaagatctccatcccggtgtaaccggtat	368
Db	387	CATCATCAACTCGATGGAAGAAACAATCTCGTGTGGAAGTATCTCTGAACTACGGTGAAT	446
QY	369	catctggactctgatcgatatcaacgcggttaagaccaaatctgtatctctcgaatacaaat	428
Db	447	CATCTGGACTCTCAGGACACTCAGGAATCAACACAGCGTGTGTATTCAAATATCTCTCA	506
QY	429	cogtgaagacatctctgaaatacatcaatcgctggttcttcgttaaccataccaataa- --	485
Db	507	GATGATCAACAATCTCTGATACATCAATCGCTGGATCTTCGTTACCATCACCAACAATCG	566
QY	486	cttgaacaatgctaaatacatcaacgcggttaaacctggaatcttaataccgacatcaaga	545
Db	567	TCGGAATAACTCCAAATCTATCAACGGCGGTCTGTATCGACCAGAAACCGATCTCCAA	626
QY	546	catcccgtaggttatcgcgttaacgcggtgaaatcatcttcaaacctggacggtgacatcgatcg	605
Db	627	TCGTGGCTAATCCACGCTTCTTAATAACATCATGTTCAAACCTGGACGGTTGTCTGTGACAC	686
QY	606	taaccagttcaatcgatgaataactcttccatcttcaacaccgcgaactgtctcagtccaa	665
Db	687	TCACCGCTACATCTGGATCAAAATCTCAATCTGTTTCGACAAAGAACTGAACGAAAAAGA	746











```
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3891 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3888
; US-08-405-496A-27

Query Match          9.2%; Score 123; DB 2; Length 3891;
Best Local Similarity 50.2%; Pred. No. 9.5e-26;
Matches 396; Conservative 0; Mismatches 375; Indels 18; Gaps 3;

Qy 20 aatacaattccgaaatcgaacataatcatctgaacctgcgttacaaagacaacaatc 79
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2603 AATATATTAAAGATATTATTAACTCTCTATATTGCAATTAAGATATGAAGTAAATCAT 2662

Qy 80 tgatgatctctcgtgggtgacggtgtaaa-----gttgaagtatacagcgggtgtgaac 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2663 TAATAGACTTATCTAGGTATGATCAAAATAAATATTGGTAGTAAAGTAAATTTTGATC 2722

Qy 134 tgaatgacaagaaccaggttcaactccttcgcctaaactcctaagatcgcgtgttaactc 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2723 CAATAGATAAAATCAAAATTCATTTATTTTAAATTTGAAAGTAGTAAATTTGAGGTAATTT 2782

Qy 194 agaatacagaacatcatcttcaactcgcgtatcctgcgtacctctctcttccctcgtgattc 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2783 TAAAAAATGCTATTCTATATATAGTATGATGAAATTTTAGTACTAGCTTTTGGATAA 2842

Qy 254 gtatccccgaaatacagaacgacgggtatccagaattatcaccaaatgaatacaccatca 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2843 GAATTCCTAAGTATTTTAAACAGTATAAGTCTAAAT-----AATGAATATACAATAA 2893

Qy 314 tcaactgcataagaataaactctggttggaagatctccatccgcgcgggttaacgcgtatcatct 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2894 TAAATGTATGAAAAAATAATTCCAGATGGAAGTAGTACTTAATTTATGTGTAATATCT 2953

Qy 374 ggactctgatgatcaacggtgagaacaaactcgtattcttccgaatacacaacatcccggtg 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2954 GGACTTTACAGGATACTCAGGAAATAAACAAGAGTAGTTTTTAAATACAGTCAAAATGA 3013

Qy 434 aagacatctctgaaatacacaatcgtcgtgttcttaccatcacccaataaa---cctga 490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3014 TTAATATATACAGATTATATAACAGATGGATTTTGTAACTATCACTAATAATAGATTAA 3073

Qy 491 acaatgctaaactctacatacaggttaactggaatcttaataccgacatcaaaagacatccc 550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3074 ATAACCTAAAAATTTATATAATGAAGATTAAATAGATCAAAAAACCAATTTCAAATTTAG 3133

Qy 551 gtgaagttatcgtcaacggtgaaatcatcttcaaaactggacggtgacatcgatcgatccc 610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3134 GTAATATTTCATGCTAGTAAATATAATAATGTTTAAATAGATGGTGTAGAGATACACATA 3193

Qy 611 agttcatctggatgaataactcttccatcttcaacaccggaactgctcagtcaccaatcgc 670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

## RESULT 11

```
US-08-915-136-27
; Sequence 27, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3891 base pairs
; TYPE: nucleic acid
```

```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3888
US-08-915-136-27

```

Query Match 9.2%; Score 123; DB 4; Length 3891;  
Best Local Similarity 50.2%; Pred. No. 9.5e-26;  
Matches 396; Conservative 0; Mismatches 375; Indels 18; Gaps 3;

```

; Patent No. 5015882
;
; GENERAL INFORMATION:
;
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUY, JIRI
;
; TITLE OF INVENTION: VACCINES, ANTIBODIES, AND TR
;
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TR
;
; TITLE OF INVENTION: INFECTIONS
;
; FILE REFERENCE: 480.19-4 (HW)
;
; CURRENT APPLICATION NUMBER: US/08/700,651B
;
; CURRENT FILING DATE: 1997-08-14
;
; EARLIER APPLICATION NUMBER: 08/415,751
;
; EARLIER FILING DATE: 1995-04-03
;
; NUMBER OF SEQ ID NOS: 15
;
; SOFTWARE: PatentIn ver. 2.0
;
; SEQ ID NO 1
;
; LENGTH: 5163
;
; TYPE: DNA
;
; ORGANISM: Cryptosporidium parvum
;
; US-08-700-651-1

```

## RESULT 12

US-08-700-651-1  
; Sequence 1, Application US/08700651B



US-08-928-361B-4

; Sequence 4, Application US/08928361B

; Patent No. 6071518

; GENERAL INFORMATION:

; APPLICANT: Petersen, Carolyn

; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,

; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS

; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM

; TITLE OF INVENTION: SPECIES INFECTIONS

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PETERS, VERNY, JONES &amp; BIKSA

; STREET: 385 Sherman Avenue, Suite 6

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94306-1840

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/928.361B

; FILING DATE: 12-SEP-1997

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/026,062

; FILING DATE: 13-SEP-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Veirny, Hana

; REGISTRATION NUMBER: 30,518

; REFERENCE/DOCKET NUMBER: 480.76-1(HV)

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-324-1677

; TELEFAX: 650-324-1678

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5163 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-928-361B-4

Query Match 5.3%; Score 70.6; DB 3; Length 5163;

Best Local Similarity 44.3%; Pred. No. 1.6e-10;

Matches 286; Conservative 0; Mismatches 359; Indels 0; Gaps 0;

Qy 269 agaagcggtatccagattacatccacaatgaatcacaccatcatcaactgcatgaaga 328

Db 574 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 633

Qy 329 ataactgtgtggaagatctccatcccggttaacccgtatcatctgtgactgtgata 388

Db 634 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 693

Qy 389 tcaacggtgaagacaaatctgtattcttcggaatacacaacatccgtgaagacatcttgaat 448

Db 694 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 753

Qy 449 acataactcggtgttcttcgttacatcaccaataaacctgaacaaatgctaaatctaca 508

Db 754 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 813

Qy 509 tcaacggttaactggaatctaatcccgacatcgaagacatccgtgaagtattcgtaacg 568

Db 814 ACAACTACAACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 873

Qy 569 gtgaataatcttcaactgacggtgacatcgatcgatccacggttcacttggatgaat 628

Db 874 ACTACAACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 933

Qy 629 acttctccatctcaacaccggaactgtctcagtcgaataatcgaaagacggtacaagatcc 688

Db 934 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 993

Qy 689 agttacttcogaatactctgaaagactcttggggttaatccgctgtatgtacacaagaat 748

Db 994 ACAACTACAACCTACCAAGAAACCAACAACCTACTACTACTACTACTACTACTACTACT 1053

Qy 749 actatatgttcaatgtgtaacaagaactcttatacaactcaaaactgaagaagaactctccgg 808

Db 1054 ACTACTACTACCACAACAACAACAACCTACTACTACTACTACTACTACTACTACTACT 1113

Qy 809 ttggtgaaatctctgactgttccaaatacaacacagaactcttaatacatcaactccgcg 868

Db 1114 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1173

Qy 869 acctgtacatcggtgaaagttcatctatccgtcgcgaaatcttaact 913

Db 1174 ACCAAGAAACCAACAACAACCTACCACTGCCACAACAACAACCTACT 1218

RESULT 14

US-08-700-651-2

; Sequence 2, Application US/08700651B

; Patent No. 6015882

; GENERAL INFORMATION:

; APPLICANT: PETERSEN, CAROLYN

; APPLICANT: LEECH, JAMES

; APPLICANT: NELSON, RICHARD, C.

; APPLICANT: GUT, JIRI

; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS

; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF CRYPTOSPORIDIUM PARVUM

; TITLE OF INVENTION: INFECTIONS

; FILE REFERENCE: 480.19-4(HV)

; CURRENT APPLICATION NUMBER: US/08/700,651B

; CURRENT FILING DATE: 1997-08-14

; EARLIER APPLICATION NUMBER: 08/415,751

; EARLIER FILING DATE: 1995-04-03

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn ver. 2.0

; SEQ ID NO 2

; LENGTH: 5318

; TYPE: DNA

; ORGANISM: Cryptosporidium parvum

US-08-700-651-2

Query Match

Best Local Similarity 5.3%; Score 70.6; DB 3; Length 5318;

Matches 286; Conservative 0; Mismatches 359; Indels 0; Gaps 0;



OM of: US-09-910-186A-8 to: Issued\_Patents\_NA:\* out\_format : pfs  
Date: Sep 2, 2002 4:01 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=framet\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPro\_spool/US09910186/runat\_29082002\_134725\_16013/app\_query.fasta\_1.505  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62  
-TRANS=human40.cdd -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09910186.@CGL1.1.51 -NCPU=6 -ICPU=3 -LONGLOG.  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-910-186A-8  
Query length: 440  
Database: Issued Patents NA:\*  
Database sequences: 383533  
Database length: 122816752  
Search time (sec): 47.780000

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-480-604A-22		831.50	1619.44	1.6e-82	1330
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-405-496A-22		831.50	1619.44	1.6e-82	1330
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-915-136-22		831.50	1619.44	1.6e-82	1330
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-480-604A-25		831.50	1618.83	1.8e-82	1402
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-405-496A-25		831.50	1618.83	1.8e-82	1402
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-915-136-25		831.50	1607.12	7.9e-82	3891
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-405-496A-27		831.50	1607.12	7.9e-82	3891
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-915-136-27		831.50	1607.12	7.9e-82	3891
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-618-312A-1		595.00	1150.30	2.2e-56	1359
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-280-328-1		595.00	1150.30	2.2e-56	1359
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-618-312A-3		594.00	1148.31	2.8e-56	1359
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-280-328-3		594.00	1148.31	2.8e-56	1359
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-668-381A-6		594.00	1144.73	4.5e-56	1858
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-110-786A-7		575.00	1110.64	3.6e-54	1359
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-446-855A-1		163.00	272.21	1.8e-07	8920
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-150-741-1		163.00	272.21	1.8e-07	8920
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-316-083-2		137.50	242.65	7.9e-06	1431
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-276-452A-24		128.00	221.90	0.0001	1690
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-798-744-24		128.00	221.90	0.0001	1690
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-568-459A-3		126.00	206.69	0.0008	4507
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-487-826B-3		126.00	206.69	0.0008	4507
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-686-968C-1		124.50	207.77	0.0007	3164
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-480-604A-9		123.50	196.52	0.0029	7101
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-405-496A-9		123.50	196.52	0.0029	7101
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-915-136-9		123.50	196.52	0.0029	7101
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-323-170B-1		123.50	193.01	0.0046	9636
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-954-441-1		123.50	193.01	0.0046	9636
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-853-659A-8		123.50	187.55	0.0093	15512
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-853-659A-63		123.50	187.55	0.0093	15512
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-853-659A-66		123.50	187.55	0.0093	15512
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-853-659A-2		123.50	182.22	0.0184	24701
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-853-659A-3		123.50	182.22	0.0184	24701
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-853-659A-60		123.50	182.22	0.0184	24701
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-853-659A-61		123.50	182.22	0.0184	24701
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-413-118-117		120.50	206.57	0.0008	1760
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-473-446-117		120.50	206.57	0.0008	1760
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-480-604A-5		120.00	188.02	0.0087	8133
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-405-496A-5		120.00	188.02	0.0087	8133

/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-08-915-136-5 + 120.00 188.02 0.0087 8133  
/cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-951-871-1 + 117.00 194.56 0.0038 2739  
/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-851-843A-1 + 117.00 192.49 0.0049 3279  
/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-08-974-349A-109 + 117.00 192.49 0.0049 327  
/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-08-854-050-1 + 117.00 192.49 0.0049 3279  
seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-480-604A-22  
seq\_documentation\_block:  
; Sequence 22, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-01763  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1330 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1314  
US-08-480-604A-22

[illegible]

```

; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1330 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1314
; US-08-405-496a-22

```

```

alignment_scores:
  Quality: 831.50      Length: 450
  Ratio: 2.607         Gaps: 10
  Percent Similarity: 70.889   Percent Identity: 40.000

```

## alignment\_block:

US-09-910-186A-8 x US-08-405-496A-22

Align seg 1/1 to: US-08-405-496A-22 from: 1 to: 1330

```

4  LysTyrAsnSerGluLeuAsnAsnIleIleLeuAsnLeuArgTyr 20
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
28 GAATACATCAAGAACATCATCACTCCATCCATCGCTGCGGTACGA 77
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
20 SASpAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluVal 37
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
78 ATCCAATCACCTGATCGACTGCTCGCTACGCTTCCAAATCAACATCG 127
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
37 YrAspGlyValGluLeuAsn.....AspLysAsnGlnPheLysLeuThr 51
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
128 GTTCTAAAGTTACTTCGATCCCATCGACAGAAATCAGATCCAGCTGTC 177
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
52 SerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIle 68
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
178 AATCTGGAATCTTCCAAATCGAAGTATCTCGAAGATGCTATCGTATA 227
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
68 eAsnSerValPheLeuAspPheSerValSerPheThrPheArgIlePro 85
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
228 CAACTCTATGTAGGAAACTTCTCCACCTCTCTGATCCGATCCGATCC 277
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
85 YstYrLysAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIle 101
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
278 AATACCTCAAC.....TCCATCTCTGAAACATGAAATACACCATC 318
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
102 IleAsnCysMetLysAsnAsnSerGlyTyrLysIleSerIleArgGlyAs 118
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
319 ATCAACTGCATGAGAAACAAATCTGTTGGAAAGTATCTCTGAACATAG 368
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
118 nArgIleIleThrLeuIleAspLeuAsnGlyLysThrLysSerValP 135
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
369 TGAATCATCTGGAATCTCGAGGACATCAGGAAATCAAAACAGCGGTG 418

```

```

135 hePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrp 151
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
419 TATTCAAAATACTCTCAGATCATCAACATCTCTGACTACATCAATCGCTGG 468
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
152 PhePheValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrIle 167
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
469 ATCTTCGTTTACCATCAACCAAAATCGTCTGAATAAATCCAAAATCTACAT 518
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
519 CAACGCCGCTGATCGACCAAGAACGATCTCCAATCTGGTAAACATCC 568
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
569 ACGTTCTTAATAACATCATGTTCAAACTGCGCGGTTGCTCGTACACTCAC 618
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerG 217
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
619 CGCTACATCTGGATCAAAATCTTCAATCTGTCGACAAAGAACTGAACGA 668
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
217 nSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeu 234
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
669 AAAAGAAATCAAGACCTGTACGACCAACAGTCCATCTGGTATCTCTGA 718
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
234 yAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
719 AAGACTTCTGGGTGACTACCTCGAGTACGACAAACCGTACTACATGCTG 768
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLys..... 263
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
769 AATCTGTACGATCCGAACAAATACGTTGACGTCAACAAATGATGATCCG 818
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
264 .....AspSerProValGlyGluIleLeuThrArgSerL 275
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
819 CGGTTATACGTACTCGAAAGTCCGCGTCTGTTATGACTACCAACA 868
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
275 ystYrAsnGlnAsnSerLysTyrIleAsnTyrArgAspLeuTyrIleGly 291
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
869 TCTACCTGAACCTCTTCC.....CTGTACCGTGGT 897
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
292 GluLysPheIleIleArgArgLysSerAsnSerGlnSerIleAsnAspAs 308
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
898 ACCAAATTCATCATCAAGAAATACGCGTCTGGTAAC.....AAGGACAA 941
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
308 pIleValArgLysGluAspTyrIleTyrLeuAspPhePheAsnLeuAsnG 325
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
942 TATCGTTCGCAACATGATCGTGTATACATCAATGTTGTAGTTAAGAA 991
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
325 InGluTrpArgValTyrThrTyrLysTyrPheLysLysGluGluLys 341
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
992 AAGAATACCGTCTGGCTACCAATGCTCTCAGGCTGGTGTAGAAAAGATC 1041
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
342 LeuPheLeuAlaProIleSerAspSerAspGluLeuTyrAsnThrIleG 358
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1042 TTCTGCTCTCGAAATCCCGGACGTTGGTAATCTGCTCAGGTAGTGT 1091
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
358 nIleLysGluTyrAspGluGlnPro...ThrTyrSerCysGlnLeuLeu 374
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1092 AATGAATCCAGAACGACGAGGATATCACTAAACAATGCAAAATG.... 1137
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
374 helYsLysAspGluGluSerThrAspGluIleGlyLeuIleGlyIleHis 390
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1138 ..AATCTGCAGGACAAACATGGTAACGATATCGTTTCATCGGTTCCAC 1185
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
391 ArgPheTyrGluSerGlyIleValPheGluGluTyrLysAspTyrPheCy 407
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1186 CAGTTCAACAATATCGCT.....AAACTGGT 1211
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
407 sIleSerLysTyrTrpTyrLeuLysGluValLysArgLysProTyrAsnLeu 424
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1212 TGGTTCCAACTGGTACAATCGTCAGATCGAACGTTCC.....TCTCGCA 1255
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
424 ysLeuGlyCysAsnTrpGlnPheIleProLysAspGluGlyTrpThrGlu 440

```





```

1258 CAGTTCACAAATATCGCT.....AAACTGGT 1280
407 fileSerLystrPrtyrLeuLysGluValLysArgLysProTyAsnLeuL 424
||||| : : : : : ||||| : : : : : ||||| : : : : :
1284 TCGTTCCCACTGGTACAATCGTCAGATCAGACGTTC .....TCTCGCA 1327
424 ysLeuGlyCysAsnTrpGlnPheIleProLysAspGluGlyTrpThrGlu 440
||||| : : : : : ||||| : : : : : ||||| : : : : : |||
1328 CTCTGGTGTCCTCTGGGAGTTCATCCCGTTGATGACGTTGGGTGA 1377

seq_name: /cgn2_6/pdodata/2/ina/5B_COMB.seq:US-08-405-496A-25

seq_documentation_block:
; Sequence 25, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405.496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1402 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1386
; US-08-405-496A-25

alignment_scores:
Quality: 831.50 Length: 450
Ratio: 2.607 Gaps: 10
Percent Similarity: 70.889 Percent Identity: 40.000

```

alignment_scores:		
Quality:	831.50	Length: 450
Ratio:	2.607	Gaps: 10
Percent Similarity:	70.889	Percent Identity: 40.000



alignment\_block:

US-09-910-186A-8 x US-08-405-496A-25 ..

Align seg 1/1 to: US-08-405-496A-25 from: 1 to: 1402

```
4  LysTyrAsnSerGluLeuLeuAsnAsnIleIleLeuAsnLeuArgTyrLy 20
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
100 GAATACATCAAGAACAATCATCAATACCTCCATCTGAACCTGGCTACGA 149
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
20  SASPAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluValt 37
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
150 ATCCAATCACCTGATCGACTGCTCGCTACGCTTCCAAAATCAACATCG 199
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
37  yRAspGlyValGluLeuAsn.....AspLysAsnGlnPheLysLeuThr 51
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
200 GTTCTAAAGTTAACTTCGATCCGACACAAGAAATCAGATCCAGCTGTT 249
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
52  SerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePh 68
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
250 AATCTGGAATCTTCCAAAATCGAAGTTATCTCTGAAGAATGCTATCGTAT 299
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
68  eAsnSerValPheLeuAspPheSerValSerPheThrIleArgIleProL 85
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
300 CAACTCTATGTAGAAACTTCTCCACCTCCTTCTGGATCCGATCCCGA 349
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
85  ySTyrlYsAsnAspGlyIleGlnAsnTyrlIleHisAsnGluTyrlIle 101
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
350 AATACITCAAC.....TCCATCTCTGAAACAATGAATACACCATC 390
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
102 IleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAs 118
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
391 ATCAACTGATGTAAGAAACAATCTGGTTGAAAGATATCTCTGAACATACG 440
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
118 nArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValP 135
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
441 TGAATATCATCTGGACACTCTGAGGAACTCAGGAAATCAACACGCTGTG 490
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
135 hePheGluTyrlYsAsnIleArgGluAspIleSerGluTyrlIleAsnArgTrp 151
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
491 TATTCAAAATACTCTCAGATGATCAACATCTCTGACTACATCAATCGCTG 540
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
152 PhePheValThrIleThrAsnAsn.....LeuAsnAsnAlaLysIleTyrl 167
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
541 ATCTCTGTTACCATCAACCAATCTCTGTAATCAATCCAAAATCTACAT 590
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
591 CAACGCCGCTCTGATCGACGAAACCGGATCTCCAATCTGGGTAACATCC 640
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
641 ACGCTTCTAATACATCATGTTCAAACTGGACGGTGTGCTGACACTCAC 690
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGI 217
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
691 CGTACATCTGGATCAAAATCAATCTCAATCTGTCGACAAAAGAACTGAAC 740
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
217 nSerAsnIleGluGluArgTyrLysIleGlnSerTyrlSerGluTyrlLeu 234
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
741 AAAAGAAATCAAGAGCCTGTACGACCAACCACTCCAATCTGGTATCTCT 790
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
234 yAspPheTrpGlyAsnProLeuMetTyrlAsnLysGluTyrlTyrlMetPhe 250
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
791 AAGACTCTGGGGTGACTACCTGCATGACGACAAACCGTACTACATGCTG 840
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
251 AsnAlaGlyAsnLysAsnSerTyrlIleLysLeuLysLys..... 263
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
841 AATCTGTACGATCCGAACAATACGTTGACGTCACAAATAGTAGTATCCG 890
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
264 .....AspSerProValGlyGluIleLeuThrArgSerL 275
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
891 CGGTTACATGTACCTGAAAGGTCGCGGTCTCTGTATGACTACCAACA 940
```

```
275 ySTyrlAsnGlnAsnSerLysTyrlIleAsnTyrlArgAspLeuTyrlIleGly 291
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
941 TCTACCTGAACCTCTCC.....CTGTACCGTGGT 969
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
292 GluLysPheIleIleArgLysSerAsnSerGlnSerIleAsnAspAs 308
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
970 ACCAAATTCATCATCAAGAAATACGGCTGTGGTAAC.....AAGGACAA 1013
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
308 pileValArgLysGluAspTyrlIleTyrlIleuAspPhePheAsnLeuAsnG 325
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1014 TATCGTTTCGCAACAATGCGTGTATACATCAATGTGTAGTTAAGAAC 1063
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
325 lngLutrpArgValTyrlTyrlLysTyrlPheLysLysGluGluGluLys 341
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1064 AAGAATACGCTCTGGCTACCAATGCTTCTCAGCGCTGTAGAAAAGATC 1113
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
342 LeuPheLeuAlaProIleSerAspSerAspGluLeuTyrlAsnThrIleGI 358
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1114 TTGTCTGCTCTGGAAATCCGGACGTTGGTAATCTCTCTCAGGTAGTTGT 1163
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
358 nIleLysGluTyrlAspGluGlnPro...ThrTyrlSerCysGlnLeuLeup 374
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1164 AATGAATCCAGACGACCGGATATCACTAACAAATGCAAAATG.... 1209
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
374 helYsLysAspGluGluSerThrAspGluIleGlyLeuIleGlyIleHis 390
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1210 ..AATCTGCAGGACACAATGTTAAGATATCGTTTCATCGGTTCCAC 1257
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
391 ArgPheTyrlGluSerGlyIleValPheGluGluTyrlLysAspTyrlPheCy 407
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1258 CAGTTCAACAATATCTGCT.....AACTGGT 1283
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
407 sIleSerLysTyrlTyrlLysGluValLysArgLysProTyrlAsnLeuL 424
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1284 TCGTTCCAACTGGTACAAATCGTCAGATCGAACGTTCC.....TCGCGCA 1327
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
424 ysLeuGlyCysAsnTrpGlnPheIleProLysAspGluGlyTrpThrGlu 440
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1328 CTCTGGTTGCTCTTGGGATTCATCCCGTTGATGACGGTTGGGTGAA 1377
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-915-136-25
seq_documentation_block:
; Sequence 25, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1402 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1386
; US-08-915-136-25

```

```

alignment_scores:
  Quality: 831.50      Length: 450
  Ratio: 2.607        Gaps: 10
Percent Similarity: 70.889 Percent Identity: 40.000

```

alignment\_block:

US-09-910-186a-8 x US-08-915-136-25 ..

Align seg 1/1 to: US-08-915-136-25 from: 1 to: 1402

```

4 LysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTyrLy 20
   ::::::::::::::::::::|
100 GAATACATCAAGAACATCATCAATACCTCCATCCTGGAACCTGCGCTACGA 149
   ::::::::::::::::::::|
20 sAspAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluValT 37
   ::::::::::::::::::::|
150 ATCCAATCACCTGATCGACCTGCTCGCTACGCTCCCAAAATCAACATCG 199
   ::::::::::::::::::::|
37 yrAspGlyValGluLeuAsn.....AspLysAsnGlnPheLysLeuThr 51
   ::::::::::::::::::::|
200 GTTCTAAAGTTAACTTCGATCCGATCCGACAAAGAATCAGATCCAGCTGTC 249
   ::::::::::::::::::::|
52 SerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePh 68
   ::::::::::::::::::::|
250 AATCTGGGAATCTTCCAATAATCGAAGTTATCTCTGAAGAATGCTATCGTATA 299
   ::::::::::::::::::::|
68 eAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleProL 85
   ::::::::::::::::::::|
300 CAACCTCTATGTACGAAACTTCTCCACCTCTCTGGATCGGTATCCCGA 349
   ::::::::::::::::::::|
85 ysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIle 101
   ::::::::::::::::::::|
350 AATACTTCAAC.....TCCATCTCTCGAACAATGAATAACACCATC 390
   ::::::::::::::::::::|
102 IleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAs 118
   ::::::::::::::::::::|
391 ATCAACTGCATGGAAACAATTCGTGGTTGAAGATATCTCTGAACATACGG 440

```

```

118 nArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValP 135
   ::::::::::::::::::::|
441 TGAATCATCTGGACTTCGAGGACACTCAGGAAATCAACAGCGTGTG 490
   ::::::::::::::::::::|
135 hePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrp 151
   ::::::::::::::::::::|
491 TATTCAAAATACTCTCAGATGATCAACATCTCTGACTATCATCAATCGCTGG 540
   ::::::::::::::::::::|
152 PhePheValThrIleThrAsnAsn....LeuAsnAsnAlaLysIleTyrIl 167
   ::::::::::::::::::::|
541 ATCTTCGTTACCATCACCACAACTCGTCTGAATACTCCAAAATCTACAT 590
   ::::::::::::::::::::|
167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184
   ::::::::::::::::::::|
591 CAACGGCGCTCTGATCGACAGAACCGATCTCCAATCTGGGTAAACATCC 640
   ::::::::::::::::::::|
184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
   ::::::::::::::::::::|
641 ACGCTTCTTAATAACATCATGTTCAAACTGGACGGTTGTCGTGACACTCAC 690
   ::::::::::::::::::::|
201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGl 217
   ::::::::::::::::::::|
691 CGCTACATCTGGATCAATACTTCAATCTGTTCGACAAAGAACTGAACGA 740
   ::::::::::::::::::::|
217 nSerAsnIleGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234
   ::::::::::::::::::::|
741 AAAGAAATCAAGACCTGTACGACAAACCCAGTCCAATTTCTGTATCCCTGA 790
   ::::::::::::::::::::|
234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
   ::::::::::::::::::::|
791 AAGACTTCTGGGTGACTACCTGCTGACGACGACAAACCCGTACTACATGCTG 840
   ::::::::::::::::::::|
251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLys..... 263
   ::::::::::::::::::::|
841 AATCTGTACGTCGACAAATACTGTGACGTCAACAATGTAGGTATCCG 890
   ::::::::::::::::::::|
264 .....AspSerProValGlyGluIleLeuThrArgSerL 275
   ::::::::::::::::::::|
891 CGGTTACATGTACCTGAAAGCTCCGCGTGTCTGTTATGACTACCAACA 940
   ::::::::::::::::::::|
275 ysTyrAsnGlnAsnSerLysTyrIleAsnTyrArgAspLeuTyrIleGly 291
   ::::::::::::::::::::|
941 TCTACCTGAACCTCTCC.....CTGTACCGTGGT 969
   ::::::::::::::::::::|
292 GluLysPheIleIleArgArgLysSerAsnSerGlnSerIleAsnAspAs 308
   ::::::::::::::::::::|
970 ACCAAATCATCATCAGAAATACCGTCTGCTGTAAC.....AAGACAA 1013
   ::::::::::::::::::::|
308 pIleValArgLysGluAspTyrIleTyrLeuAspPhePheAsnLeuAsnG 325
   ::::::::::::::::::::|
1014 TATCGTTCCGCAACAATGATCGTGATATACATCAATGTTGTAGTTAAGAACA 1063
   ::::::::::::::::::::|
325 lnGluTrpArgValTyrThrTyrLysTyrPheLysLysGluGluLulys 341
   ::::::::::::::::::::|
1064 AAGAATACCGTCTGGCTACCAATGCTTCTCAGGCTGGGTAGAAAAGATC 1113
   ::::::::::::::::::::|
342 LeuPheLeuAlaProIleSerAspSerAspGluLeuTyrAsnThrIleGl 358
   ::::::::::::::::::::|
1114 TTGTCGCTCTCGGAATCCCGGACGTTGGTAATCTGCTCAGGTAGTTGT 1163
   ::::::::::::::::::::|
358 nIleLysGluTyrAspGluGlnPro...ThrTyrSerCysGlnLeuLeup 374
   ::::::::::::::::::::|
1164 AATGAAATCCAAAGAACGACCGAGGTATCAGTAAACAAATGCAAAATG.... 1209
   ::::::::::::::::::::|
374 heLysLysAspGluSerThrAspGluIleGlyLeuIleGlyIleHis 390
   ::::::::::::::::::::|
1210 ..AATCTCAGGACAACTGTTACGATATCGGTTTCATCGGTTTCCAC 1257
   ::::::::::::::::::::|
391 ArgPheTyrGluSerGlyIleValPheGluGluTyrLysAspTyrPheCy 407
   ::::::::::::::::::::|
1258 CAGTTCAACAATATCGCT.....AACTGGT 1283
   ::::::::::::::::::::|
407 sIleSerLysTrpTyrLeuLysGluValLysArgLysProTyrAsnLeuL 424

```

||||:||||| :||||:||||| :|||  
1284 TGCTTCCAAGTGTACAACTGCTCAGATCGAAGCTTC .....TCTCGCA 1327

424 ysLeuGlyCysAsnTrpGluPheIleProLysAspGluGlyTrpThrGlu 440  
||||:||||| :||||:||||| :|||  
1328 CTCGGGTGCTCTGGAGATTCAATCCGTTGATGACGGTGGGGTAA 1377

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:us-08-480-604A-27

seq\_documentation\_block:

; Sequence 27, Application US/08480604A

; Patent No. 5736139

; GENERAL INFORMATION:

; APPLICANT: KINK, JOHN A.

; APPLICANT: THALLEY, BRUCE S.

; APPLICANT: PADHYE, NISHA V.

; APPLICANT: FIRCA, JOSEPH R.

; APPLICANT: STAFFORD, DOUGLAS C.

; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND

; PREVENTION OF C. DIFFICILE DISEASE

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MEDLEN & CARROLL, LLP

; STREET: 220 MONTGOMERY STREET, SUITE 2200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/480,604A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/422,711

; FILING DATE: 14-APR-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/405,496

; FILING DATE: 16-MAR-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/329,154

; FILING DATE: 25-OCT-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/161,907

; FILING DATE: 02-DEC-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/985,321

; FILING DATE: 04-DEC-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/429,791

; FILING DATE: 31-OCT-1989

; ATTORNEY/AGENT INFORMATION:

; NAME: INGOLIA, DIANE E.

; REGISTRATION NUMBER: 40,027

; REFERENCE/DOCKET NUMBER: OPHD-01763

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 705-8410

; TELEFAX: (415) 397-8338

; INFORMATION FOR SEQ ID NO: 27:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3891 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..3888

US-08-480-604A-27

alignment\_scores:

Quality: 831.50 Length: 450

Ratio: 2.607 Gaps: 10

Percent Similarity: 70.889 Percent Identity: 40.000

alignment\_block:

US-09-910-186A-8 x US-08-480-604A-27 ..

Align seg 1/1 to: US-08-480-604A-27 from: 1 to: 3891

4 LysTrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTyrLy 20  
:||||:||||| :||||:||||| :|||  
2602 GAATATATTAAAGATATTATTAACTCTCTATATTGAATTAAGATATGA 2651

20 sAspAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluValt 37  
:||||:||||| :||||:||||| :|||  
2652 AAGTAATCAATTAATAGACTTATCTAGGTATGCATCAAAATAAATATTG 2701

37 yrAspGlyValGluLeuAsn.....AspLysAsnGlnPheLysLeuThr 51  
:||||:||||| :||||:||||| :|||  
2702 GTAGTAAAGTAAATTTTGTCCCAATAGATAAAATCAAAATCAATATT 2751

52 SerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePh 68  
:||||:||||| :||||:||||| :|||  
2752 AATTTAGAAAGTAGTAAATTTAGGTAATTTTAAAAAATGCTATTGTATA 2801

68 eAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleProL 85  
:||||:||||| :||||:||||| :|||  
2802 TAATAGTATGTATGAAATTTTGTAGTACTAGCTTTTGGATAAGAAATTCCTA 2851

85 ysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIle 101  
:||||:||||| :||||:||||| :|||  
2852 AGTATTTTAAAC.....AGTATAAGTCTAAATAATGAATATACATA 2892

102 IleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAs 118  
:||||:||||| :||||:||||| :|||  
2893 ATAAATGTATGGAATAATTTTCTAGGATGGAAGATCACTTAATATTATGG 2942

118 nArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValP 135  
:||||:||||| :||||:||||| :|||  
2943 TGAATAATCTGGACTTTACAGGATACCTCAGAAATAAAACAAGACGTAG 2992

135 hePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrp 151  
:||||:||||| :||||:||||| :|||  
2993 TTTTAAATACAGTCAAAATGATTAAATATATACAGATTATATAAACAGATGG 3042

152 PhePheValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrIl 167  
:||||:||||| :||||:||||| :|||  
3043 ATTTTGTAACTATCACTAATAATAGATTAAATAACTCTAAATTTATAT 3092

167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184  
:||||:||||| :||||:||||| :|||  
3093 AAATGGAAGATTAAATAGATCAAAACCAATTTCAAAATTTAGGTAATATTC 3142

184 leaIaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200  
:||||:||||| :||||:||||| :|||  
3143 ATGCTAGTAATAATAATGTTTAAATTTAGATGTTTGTAGAGATACACAT 3192

201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGl 217  
:||||:||||| :||||:||||| :|||  
3193 AGATATATTGGATAAAATAATTTTAACTCTTTTTCATAGGAATTAATGA 3242

217 nSerAsnIleGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234  
:||||:||||| :||||:||||| :|||  
3243 AAAAGAAATCAAGATTTTATATGATATCAATCAATCAATCAAGTATTATA 3292

234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250  
:||||:||||| :||||:||||| :|||  
3293 AAGACTTTTGGGTGATTATTACAAATATGATAAACCATACTATATGTATA 3342

251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLys..... 263

```

|||
3343 AATTATATGATCCAAATAAATATCGATGTAATAATGATAGGTATAG 3392
264 .....AspSerProValGlyGluLeuLeuThrArgSerL 275
3393 AGTTATATGATCTTAAGGGCTAGAGGTAGCGTAATGACTACAAACA 3442
275 ystYrAsnGlnAsnSerLysTyrIleAsnTyrArgAspLeuTyrIleGly 291
3443 TTATTATTAATTCAGT .....TTGTATAGGGGG 3471
292 GluLysPheIleIleArgLysSerAsnSerGlnSerIleAsnAspAs 308
3472 ACAAAATTTATTAATAAAATATGCTTCGCAAT .....AAGATAA 3515
308 pIleValArgLysGluAspTyrIleTyrLeuAspPheAsnLeuAsnG 325
3516 TATTGTTAGAAATAATGATCGTGTATATATTAGTAGTAAATAATA 3565
325 InGluTyrArgValTyrThrTyrLysTyrPheLysLysGluGluLys 341
3566 AAGATATAGTTAGTCTACTAATGATCATCACAGGCGGTAGAAATAATA 3615
342 LeuPheLeuAlaProIleSerAspSerAspGluLeuTyrAsnThrIleG 358
3616 CTAAGTGCAATTAGAAATACCTGATAGGAAATCTAAGTCAAGTAGTAGT 3665
358 nIleLysGluTyrAspGluGlnPro...ThrTyrSerCysGlnLeuLeuP 374
3666 AATGAAGTCAAAATAATGATCAGGAAATAACAATAAATGCAAAATG... 3711
374 heLysLysAspGluSerThrAspGluIleGlyLeuIleGlyIleHis 390
3712 ..AATTCAAGATATATGGAATGATATAGGCTTATAGGATTTCAT 3759
391 ArgPheTyrGluSerGlyIleValPheGluGluTyrLysAspTyrPheCy 407
3760 CAGTTTAATAATATAGT .....AAACTAGT 3785
407 stIleSerLysTyrTyrLeuLysGluValLysArgLysProTyrAsnLeu 424
3786 AGCAAGTAATGGTATAATAGCAAAATAGAAATCT .....AGTAGGA 3829
424 ysLeuGlyCysAsnTyrGlnPheIleProLysAspGluGlyTyrThrGlu 440
3830 CTTTGGTGTCTCAGGAAATTTCTCTGATGATGATGGATGGGAGNA 3879
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-405-496A-27
seq_documentation_block:
Sequence 27, Application US/08405496A
Patent No. 5919665
GENERAL INFORMATION:
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
COMPUTER: IBM PC compatible
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 3891 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3888
US-08-405-496A-27

alignment_scores:
Quality: 831.50 Length: 450
Ratio: 2.607 Gaps: 10
Percent Similarity: 70.889 Percent Identity: 40.000

alignment_block:
US-09-910-186A-8 x US-08-405-496A-27 ..
Align seg 1/1 to: US-08-405-496A-27 from: 1 to: 3891
4 LysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTyrly 20
2602 GAATATATAAGAAATATTAATACTTCTATATTAAGATTAAGATGA 2651
20 sasAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluValT 37
2652 AAGTAATCATTTAATAGACTTATCTAGGTATGATCAAAAATAAATAT 2701
37 yraspGlyValGluLeuAsn.....AspLysAsnGlnPheLysLeuThr 51
2702 GTAGTAAGTAATAATTTTATCAATAGATAAAATCAAAATCAATATTT 2751
52 SerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePh 68
2752 AATTTAGAAAGTAGTAAATTCAGGCTAATTTTAAAAAATGCTATTGTATA 2801
68 eAsnSerValPheLeuAspPheSerValSerPheThrPleArgIleProL 85
2802 TAATAGTATGTAGAAATTTTATGACTAGCTTTTGGATAAGAAATTCCTA 2851
85 ystYrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIle 101
2852 AGTATTTTAAC.....AGTATAAGTCTAAATAATGAATATACAAATA 2892
102 IleAsnCysMetLysAsnAsnSerGlyTyrPlyIleSerIleArgGlyAs 118
2893 ATAAATTTGATGGAATAAATTCAGGATGGAAGATCATCATTAAATATGG 2942
118 nArgIleIleThrPheLeuIleAspIleAsnGlyLysThrLysSerValP 135
2943 TGAATAATCTGGACTTTCACAGGATACTCAGGAAATAAACAAGAGTAG 2992
```

```
135 hePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrp 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2993 TTTTAAATACAGTCAAAATGATTAAATATATACAGATTATATAACAGATGG 3042
152 PhePheValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrIle 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3043 ATTTTGTAACTACTACTATAATAGATTAAATAAATACTTAAATTTATAT 3092
167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIle 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3093 AAATGAAGATTAAATAGATCAAAACCAATTTCAATTTAGTTAAATATTC 3142
184 leaIaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3143 ATCTAGTAAATAATAATGTTTAAATTTAGATGGTTGTAGAGATACACAT 3192
201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGI 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3193 AGATATATTTGGATAAAATATTTTAAATCTTTTGTAAAGAAATTAATGA 3242
217 nSerAsnIleGluIuArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3243 AAAGAATAACAAGATTATATGATATCAATCAUAAATTCAGGTATTTTAA 3292
234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3293 AAGACTTTTGGGTGATTATTTACAATATGATAAACCATACATATATGTGA 3342
251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLys..... 263
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3343 AATTATATGATCAATCAAAATATATTCGATGTAATAATGTTAGTATTAG 3392
264 .....AspSerProValGlyGluIleLeuThrArgSerL 275
3393 AGGTTATATGATCTTAAAGGGCTAGAGGTAGCGTAATGACTACAAACA 3442
275 ystYrAsnGlnAsnSerLysTyrIleAsnTyrArgAspLeuTyrIleGly 291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3443 TTTTATTAAATTCACCT.....TTGTATAGGGGG 3471
292 GluLysPheIleIleArgLysSerAsnSerGlnSerIleAsnAspAs 308
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3472 ACAAAATTTTATATAAAAAATATGCTTCTGGAAT.....AAAGATAA 3515
308 pIleValArgLysGluAspTyrIleTyrLeuAspPheAsnLeuAsnG 325
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3516 TATGTAGAAATAATGCTGATATATATTAATGTAGTAGTTAAATAA 3565
325 InGluTrpArgValTyrThrTyrLysTyrPheLysLysGluGluLys 341
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3566 AAGAATATAGTTAGTACTACTAATGCATCACAGCGCAGCGTAGAAAAATA 3615
342 LeuPheLeuAlaProIleSerAspSerAspGluLeuTyrAsnThrIleGI 358
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3616 CTAAGTGCATTAGAAATACCTCTAGTAAATCTAAGTCAAGTAGTAGT 3665
358 nIleLysGluTyrAspGluGlnPro...ThrTyrSerCysGlnLeuLeuP 374
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3666 ATGAACTCAAAAATGATCAGGAATAACAATAATCAATGCAAAATG.... 3711
374 heLysLysAspGluGluSerThrAspGluIleGlyLeuIleGlyIleHis 390
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3712 ..AATTTACAAGATAATAATGGGAATGATATAGGCTTTTATAGGATTTCA 3759
391 ArgPheTyrGluSerGlyIleValPheGluGluTyrLysAspTyrPheCy 407
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3760 CAGTTTAAATAATAGCT.....AAACTAGT 3785
407 sIleSerLysTyrTrpTyrLeuLysGluValLysArgLysProTyrAsnLeuL 424
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3786 ACCAAGTATTTGGTATATAGACAAATAAGAAATCT.....AGTAGGA 3829
```

```
424 ysLeuGlyCysAsnTrpGlnPheIleProLysAspGluGlyTrpThrGlu 440
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3830 CTTTGGTTCCTCATGGGAATTTATTCCTCTAGATGATGATGGGAGAA 3879
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-915-136-27
seq_documentation_block:
; Sequence 27, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3891 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3888
; US-08-915-136-27
alignment_scores:
```



```

; APPLICATION NUMBER: GB 8926832.0
; FILING DATE: 28-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9006097.1
; FILING DATE: 17-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Crawford Mc, Arthur R
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 510-51
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 0101 703 8750400
; TELEFAX: 0101 703 5253468
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1356
; US-07-618-312A-1

```

## alignment\_scores:

```

Quality: 595.00 Length: 441
Ratio: 2.017 Gaps: 12
Percent Similarity: 66.893 Percent Identity: 32.200

```

## alignment\_block:

```

US-09-910-186a-8 x US-07-618-312A-1 ..
Align seg 1/1 to: US-07-618-312A-1 from: 1 to: 1359

14 IleLeuAsnLeuArgTyrLysAspAsnLeuIleAspLeuSerGlyTy 30
|||||
67 ATTTAAATTTAGATATTAATGAATGATTAATATCGATATCTGGGT 116
|||||
30 rGlyAlaLysValGluValTyrAspGlyValGluLeu.....A 43
|||||
117 TAATTCATCTGTAATAACATATCCAGATGCTCAATTGGTCCCGGAATA 166
|||||
43 snAspLysAsnGlnPheLysLeuThrSerSerAlaAsnSerLysIleArg 59
|| |||
167 ATGGCAAA...GCAATACATTTTAGTAACAATGAATCTTCTGAAGTTATA 213
|||||
60 ValThrGlnAsnGlnAsnIlePheAsnSerValPheLeuAspPheSe 76
|||
214 GTGCATAAGCTATGATATGTAATGATATGATATGTTTAATAATTTTAC 263
|||||
76 rValSerPheTrpIleArgIleProLysTyrLysAsnAspGlyIleGlnA 93
|||||
264 CGTTAGCTTTTGGTTGAGGTTCTCTAAAGTATCTGCTAGTCATTTAGAAC 313
|||||
93 snTyrIleHisGlnGluTyrThrIleIleAsnCysMetLysAsnAsn... 108
|||||
314 AATATGGCAACAATGATGATTAATTAATTAATGCTCTATGAAAAACATAGT 363
|||||
109 .....SerGlyTrpLysIleSerIleArgGlyAsnArgIleI 121
|||||
364 CTATCAATAGGATCTGGTGGAGTGATCACTTAAAGGTAATAAATCAAT 413
|||||
121 etrpThrLeuIleAspIleAsnGlyLysThrLysSerValPhePheGluT 138
|||||
414 ATGACTTTAAAGATTCGCGGAGAGTATAGACAAATAAATCTTTAGG. 462
|||||
138 yrAsnIleArgGluAspIleSerGluTyrIle...AsnArgTrpPhePhe 153
|||||

```

```

463 ..GATTTACCTGATAAATTTAATGCTTTATTTAGCAAAATAAATGGTTTTT 510
154 ValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrIleAsnG 169
|||||
511 ATAACATATTACTAATGATGATGATTTCTTCTGCTAAATTTGTATATAATGG 560
|||||
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAlaA 186
|||||
561 AGTACTTATGGGAAGTCGCAAAATTTACTGGTTTAGGAGCTATTAGAGAGG 610
|||||
186 snGlyGluIleLlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|||||
611 ATAATAATATACATTAATAACTAGATAGATGTAATAATAATAATCAATAC 660
|||||
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||
661 GTTCTTATTGATAAATTTAGGATATTTTGCAAAGCATTAATAATCCAAAGA 710
|||||
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLysLysasp 236
|||||
711 GATTGAAAAATTTATACACAAGTTATTATCTATAACCTTTTAAAGAGACT 760
|||||
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||
761 TCTGGGAAACCCCTTACGATATGATACAGATATATTATTAATACCAGTA 810
|||||
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysaspSerProValGlyG 269
|||||
811 GCTTCTAGTTCTTAAGATGTTCAATTTGAAAAATATAACAGATTATATGTA 860
|||||
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||
861 TTTGACAATGCCCCATCGTATACGCGAAATTCGAATATATATATATA 910
|||||
286 rgAspLeuTyrIleGlyGluLysPheIleIleArgLysSerAsnSer 302
|||||
911 GAAGGTTATATAATGGACTAAAAATTTATATAAAAGATATACACCTAAT 960
|||||
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrIleuAs 319
|||||
961 AATGAATA...GATCTTTTGTAAATCAGGTGATTTTATTAAATTA.. 1005
|||||
319 pPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyr..... 332
|||||
1006 .TATGATCATATAACAATAAATGAGCAGCATTTGAGTTATCCGAAAGATG 1054
|||||
333 ..LysTyrPheLysLysGluGluLysLeuPheLeuAlaProIleSer 348
|||||
1055 GAAATGCTTTAATAATCTTGTAGATAATCTTAAGAGTAGGTTAATATGCC 1104
|||||
349 AspSerAspGluLeuTyrAsnThrIleGlnIleLysGluTyrAspGluG 365
|||||
1105 CAGGATCCCTCTTTATAAAAAAATGGAAGCAGTAAAAATTCGCTGATTT 1154
|||||
365 nProThrTyrSerCysGlnLeuLeuPheLysLysAspGluGluSerThrA 382
|||||
1155 AAAAACTTCTGTACAACTTAAATTTATATGATGATAAAAAATGCACTCT. 1203
|||||
382 spGluIleGlyLeuIleGlyIleHisArgPheTyrGluSerGlyIleVal 398
|||||
1204 ....TTAGGACTAGTAGGTACCCAT.....AATGTCAAATAGGC 1239
|||||
399 PheGluGluTyrLysAspTyrPheCysIleSerLysTrpTyrIleuLysG 415
|||||
1240 AAGATCCAAATAGGATATTAATTTGCAAGCACTGGTACTTTAATCA 1289
|||||
415 uValLysArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheI 432
|||||
1290 TTTAAAGATAAA.....ATTTTAGGATGATGATTGGTACTTTCG 1327
|||||
432 leProLysaspGluGlyTrpThr 439
|||||
1328 TACCTACAGATGAAGGATGGACA 1350

```

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-280-228-1

```

seq_documentation_block:
; Sequence 1, Application US/08280228
; Patent No. 5571694
; GENERAL INFORMATION:
; APPLICANT: Makoff Dr, Andrew J
; APPLICANT: Romanos Dr, Michael A
; APPLICANT: Clare Dr, Jeffrey J
; APPLICANT: Fairweather Dr, Neil F
; TITLE OF INVENTION: VACCINES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NO. 5571694th Glebe Road
; CITY: Arlington,
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/280,228
; FILING DATE: 25-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/618,312
; FILING DATE: 27-NOV-1990
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8926832.0
; FILING DATE: 28-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9006097.1
; FILING DATE: 17-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mary J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN OR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1356
; US-08-280-228-1

```

alignment\_scores:  
 Quality: 595.00 Length: 441  
 Ratio: 2.017 Gaps: 12  
 Percent Similarity: 66.893 Percent Identity: 32.200

alignment\_block:  
 US-09-910-186a-8 x US-08-280-228-1 ..

Align seg 1/1 to: US-08-280-228-1 from: 1 to: 1359

```

14 IleLeuAsnLeuA:rgTyrLysAspAsnAsnLeuIleLeuAspLeuSerGlyTy 30
|||||
67 ATTTTAAATTTAGATATTAAATGATATTAAATCGATATATCGGGT 116
30 rGlyAlaLysValGluValTyrAspGlyValGluLeu.....A 43
:|||||
117 TAATTTCATCTGTAATAACATATCAGATGCTCAATTGGTCCCGGAATAA 166
43 snAspLysAsnGlnPheLysLeuThrSerSerAlaAsnSerLysIleArg 59
|||||
167 ATGGCAAA...GCAATACATTTAGTAACAATGAATCTCTGAAGTTATA 213
60 ValThrGlnAsnGlnAsnIlePheAsnSerValPheLeuAspPheSe 76
|||
214 GTGCATAAAGCTATGGATATGAATATATGATATGTTTAAATAATTTAC 263
76 rValSerPheTrpIleArgIleProLysTyrLysAsnAspGlyIleGlnA 93
:|||||
264 CGTTAGCTTTTGGTTGAGGTTCTCTAAAGATATCTGCTAGTCATTTAGA 313
93 snTyrIleHisAsnGluTyrThrIleIleAsnCysMetLysAsnAsn... 108
:|||||
314 AATATGGCACAATAGATATCAATAATAGCTCTATGAAAAACATAGT 363
109 .....SerGlyTrpLysIleSerIleArgGlyAsnArgIleI 121
|||||
364 CTATCAATAGGATCGGTTGGAGTGATCACTTAAGGTAATAACTTAAT 413
121 eTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPhePheGluT 138
|||||
414 ATGGACTTTAAAGATTCCGCGGAGAAAGTTAGACAAATAACTTTTAGG. 462
138 yrAsnIleArgGluAspIleSerGluTyrIle...AsnArgTrpPhePhe 153
:|||||
463 ..GATTTACCTGATAAATTTAATGCTTATTTAGCAAAATAAATGGGTTT 510
154 ValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrIleAsnGl 169
:|||||
511 ATAATACTATTAATGATAGATATCTCTGCTAATTTGTATATAAATGG 560
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAla 186
|||
561 AGTACTTTATGGAGAGTCAGAAATTAAGTCTAGGAGCTATTAGAGAGG 610
186 snGlyGluIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
:|||||
611 ATAATAATAACATTAACACTAGATAGATGTAATAATAATAATCAATAC 660
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
:|||||
661 GTTCTCTATTGATAAATTTAGGATATTTTGCAAGCATTAATCCAAAGA 710
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAsp 236
:|||||
711 GATTGAAAAATTAACACAAAGTTATTTATCTATAACCTTTTAAAGAGACT 760
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||
761 TCTGGGAAACCCCTTACGATATGATACAGAAATATTTATTAACACAGTA 810
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGly 269
:|||||
811 GCTTCTAGTTCTAAAGATGTTCAATTGAAAAATATACACAGATTATATCTA 860
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
:|||||
861 TTTCACAAATCGCCCATCGTATACACTAACGGAAATTTGAATATATATA 910
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgArdLysSerAsnSer 302
|||||
911 GAAGTTATATAAAGTAAATTTATTAATAAAGATATACACCTAAT 960
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319

```



```

; : : : : ||| : : : : : : : : : : ||| : : : : |||
961 AATGAAATA...GATCTTTGTTAAATCAGGTGATTTTATTAATTA... 1005
;
319 pPhePheAsnLeuAsnGlnGluTrpArgValThrThr... 332
; : : : : ||| : : : : : : : : : : |||
1006 .TATGTATCATATAACAATAATGAGCACATGTAGGTATCCGAAAGATG 1054
;
333 .LysTyrPheLysLysGluGluLysLeuPheLeuAlaProIleSer 348
; : : : : ||| : : : : : : : : : : ||| : : : : : :
1055 GAATGCTTTAATATCTTTAGTAGAATCTTAAGAGTAGGTATTAATGCC 1104
;
349 AspSerAspGluLeuTyrAsnThrIleGlnIleLysGluTyrAspGluG1 365
; : : : : ||| : : : : : : : : : : ||| : : : : : :
1105 CCAGGTATCCCTCTTTATAAAAAAATGGAAGCACATAAAATGGCTGATTT 1154
;
365 nProThrTyrSerCysGlnLeuPheLysLysAspGluSerThrA 382
; : : : : ||| : : : : : : : : : : ||| : : : : : :
1155 AAMACCTATCTGTCACTTAATATATATGATGATAAAATGCAATCT. 1203
;
382 spGluIleGlyLeuIleGlyIleHisArgPheTyrGluSerGlyIleVal 398
; : : : : ||| : : : : : : : : : : |||
1204 .....TTAGGACTAGTAGGTACCAT.....AATGCTCAATAGGC 1239
;
399 PheGluGluTyrLysAspTyrPheCysIleSerLysTrpTyrLysGlu 415
; : : : : ||| : : : : : : : : : : ||| : : : : : :
1240 AACGATCCAAATAGGATATATTAATGCAAGCAACTGGTACTTTAATCA 1289
;
415 uValLysArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheI 432
; : : : : ||| : : : : : : : : : : ||| : : : : : :
1290 TTTAAAGATAAA.....ATTTAGGATGTGATGGTACTTTG 1327
;
432 leProLysAspGluGlyTrpThr 439
; : : : : ||| : : : : : : : : : : |||
1328 TACCTACAGATGAAGGATGACA 1350
;

```

seq\_name: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq:us-07-618-312A-3

```

seq_documentation_block:
; Sequence 3, Application US/07618312A
; Patent No. 5389540
; GENERAL INFORMATION:
; APPLICANT: Makoff Dr, Andrew J
; APPLICANT: Romanos Dr, Michael A
; APPLICANT: Clare Dr, Jeffrey J
; APPLICANT: Fairweather Dr, Neil F
; TITLE OF INVENTION: VACCINES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 14th Floor
; STREET: 2200 Clarendon Boulevard,
; CITY: Arlington,
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/618,312A
; FILING DATE: 19910516
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8926832.0
; FILING DATE: 28-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9006097.1
; FILING DATE: 17-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Crawford Mr, Arthur R
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 510-51
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 0101 703 8750400
; TELEFAX: 0101 703 5253468
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1356
; US-07-618-312A-3
;
; alignment_scores:
; Quality: 594.00 Length: 441
; Ratio: 2.014 Gaps: 12
; Percent Similarity: 66.893 Percent Identity: 31.973
;
; alignment_block:
; US-09-910-186A-8 x US-07-618-312A-3 ..
; Align seg 1/1 to: US-07-618-312A-3 from: 1 to: 1359
;
14 IleLeuAsnLeuArgTyrLysAspAsnLeuIleAspLeuSerGly 30
|||||
67 ATTCTGACCTGGACATCAACACGATATTATCTCCGACATCTCTGTTT 116
30 rGlyAlaLysValGluValTyrAspGlyValGluLeu.....A 43
|||||
117 CAACTCCTCTGTTATCATATCCAGATGCTCAATTGTCGCGGCATCA 166
43 snAspLysAsnGlnPheLysLeuThrSerSerAlaAsnSerLysIleArg 59
|||||
167 ACGGCAAA...GCTATCCACCTGGTTAAACACGAATCTTCTGAAGTTATC 213
60 ValThrGlnAsnGlnAsnIleIlePheAsnSerValPheLeuAspPheSe 76
|||||
214 GTGCACAAGCCATGGACATCGAATACACGACATGTTCAACAACCTCAC 263
76 rValSerPheTrpIleArgIleProLysTyrLysAsnAspGlyIleGlnA 93
|||||
264 CGTTAGCTTCTGCTGCGCGTTCGAAAGTTCTGCTTCCACCTGGAAC 313
93 snTyrIleHisAsnGluTyrThrIleIleAsnCysMetLysAsn... 108
|||||
314 AGTACGGCAGCAAGAGTACTCCATCATCATGCTATATGAAGAAACACTCC 363
109 .....SerglyTrpLysIleSerIleArgGlyAsnArgIleI1 121
|||||
364 CTGTCCATCGGCTCTGTTGTTGTTCTGTTTCCCTGAAGGGTAAACACCTGAT 413
121 eTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPhePheGluT 138
|||||
414 CTGACTCTGAAAGACTCCCGCGGCGAAGTTCTGTCAGATCATCTTCCGC. 462
138 yrAsnIleArgGluAspIleSerGluTyrIle...AsnArgTrpPhePhe 153
|||||
463 ..GACCTGCGCGACAGTTCACACGCTACCTGGCTGCTAAACAAATGGGTTTC 510
154 ValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrIleAsnG1 169
|||||
511 ATCATTATCATCACTAACGATCTGCTCTTCTGCTAACCTGTACATCAACGG 560
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAlaA 186
| ||| : : : : : : : : : : ||| : : : : : :

```

```

; APPLICANT: Romanos Dr, Michael A
; APPLICANT: Clare Dr, Jeffrey J
; APPLICANT: Fairweather Dr, Neil F
; TITLE OF INVENTION: VACCINES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 No. 5571694th Glebe Road
; CITY: Arlington,
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/280,228
; FILING DATE: 25-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8926832.0
; FILING DATE: 28-NOV-1989
; APPLICATION DATA:
; APPLICATION NUMBER: GB 9006097.1
; FILING DATE: 17-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mary J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1356
; US-08-280-228-3

alignment_scores:
    Quality: 594.00
    Ratio: 2.014
    Percent Similarity: 66.893    Percent Identity: 31.973

alignment_block:
    US-09-910-186A-8 x US-08-280-228-3    ..

Align seg 1/1 to: US-08-280-228-3 from: 1 to: 1359

14 tleleAsnLeuAargtyrLyAspAsnLeulleAspLeuSerGlyTy 30
|||||
67 ATCTCGAACTTGACATCAACACGATATATCTCCGACATCTCGGTTT 116
|||||
30 rclyalAtysValGluValtyrAspGlyValGluLeu.....A 43
|||||
117 CAACCTCTCTGTATACATATCCAGATGCTCAATTGTGCGGGCATCA 166
|||||

```

```
43  snAspLysAsnGlnPheLysLeuThrSerSerAlaAsnSerLysIleArg 59
||| |||
167  ACGCAAA...GCTATCCACCTGGTTAAACAACGAATCTCTGAAGTTATC 213
||| |||
60  ValThrGlnAsnGlnAsnIleIlePheAsnSerValPheLeuAspPheSe 76
||| |||
214  GTGCACAAGCCCATGGACATCGAATACAACGACATGTTCAACAACCTCAC 263
||| |||
76  rValSerPheTrpIleArgIleProLysTyrLysAsnAspGlyIleGlnA 93
:|||||
264  CGTGTAGCTTCTGGCTGGCTGCGGTTCCGAAGATTTCTGCTTCCACCTGGA 313
:|||||
93  snTyrIleHisAsnGluTyrThrIleIleAsnCysMetLysAsnAsn... 108
:|||||
314  AGTACGGCACAACGAGTACTCATCATCAGCTCTATGAAGAACAACACTCC 363
:|||||
109  .....SerGlyTrpLysIleSerIleArgGlyAsnArgIleI 121
|||||
364  CTGTCCATCGGCTCTGGTTGGTCTGTTCCCTGAAGGTAACAACCTGAT 413
|||||
121  eTrpThrLeuIleAspIleAsnGlyThrLysSerValPhePheGluTr 138
|||||
414  CTGGACTCTGAAGACTCCGCGGCGAAGTTCTGCATCAGTCACTTCCGC. 462
|||||
138  yAsnIleArgGluAspIleSerGluTyrIle...AsnArgTrpPhePhe 153
:|||||
463  ..GACCTGCGGACAGATTCAACGGTACCTGGCTAACAAATGGGTTTC 510
:|||||
154  ValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrIleAsnG 169
:|||||
511  ATCATATCATCAACGATCGTCTGCTTCTGCTTAACCTGTACATCAACGG 560
|||||
169  yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAla 186
||| |||
561  CGTCTCTGATGGGCTCGCTGAAATCTACTGCTGCGGCTATCCGTGAGG 610
|||||
186  snGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
:|||||
611  ACAACACATCATCTTAAGCTGGACGGTGGCAACACACACACACAGTAC 660
|||||
203  IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerA 219
:|||||
661  GTATCCATCGACAAGTCCGATATCTTCTGCAAAAGCACTGAACCGGAAGA 710
|||||
219  nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAsp 236
:|||||
711  GATCGAAAAAAGCTATACCAAGCTACCTGTCTATCACCTCTCTGCTGACT 760
|||||
236  heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||
761  TCTGGGGTACCCGCGGTGAGTACACCGAATATTAACCTGATCCCGGTA 810
|||||
253  GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyG 269
:|||||
811  GCCTTCTAGCTTAAGACGCTTACGCTGAAAAACATCATCACTACATGTA 860
|||||
269  uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
:|||||
861  CTTGACCAACGCGCTCTCTACTAAGCGTAAACTGAACATCTACTATACC 910
|||||
286  rGAspLeuTyrIleGlyGluLysPheIleIleArgLysSerAsnSer 302
||| |||
911  GAGCTGTGTACAAAGCGCTGAATATCATCAACAGCTACATCCCGAAC 960
|||||
303  GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319
:|||||
961  AACGAAATC...GATCTTTCGTTAAATCTGGTGACTTCACTCAAACTG... 1005
|||||
319  pPheAsnLeuAsnGlnGluTrpArgValTyrThrTyr..... 332
:|||||
1006  .TACGTTTCTTACAAACAACAAACACATCGTTGGTTACCCGCAAGACG 1054
|||||
```

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-668-381A-6

```
seq_documentation_block:
; Sequence 6, Application US/08668381A
; Patent No. 5780024
; GENERAL INFORMATION:
; APPLICANT: Brown, Robert H.
; APPLICANT: Fishman, Paul S.
; APPLICANT: Francis, Jonathan W.
; APPLICANT: Hosler, Betsy A.
; TITLE OF INVENTION: SUPEROXIDE DISMUTASE/TETANUS TOXIN
; TITLE OF INVENTION: FRAGMENT C HYBRID PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668,381A
; FILING DATE: 21-JUN-1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/000,473
; FILING DATE: 23-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,164
; REFERENCE/DOCKET NUMBER: 00786/269001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1858 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```



[illegible]

811 GCCTCTAGTCTCAAGAAGTTCACATGAAATAAATAACAAGTTATTATGA 800

269 uileLeuthrArgSerLysAsnGlnAnSerLysTyIleAsnTyra 286  
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::  
861 TTTCACAAAATGCCCATCGTACTACTAACGC AAAAATTCAAATATATATATA 910

286 rgasLeuTyrlleGlyGluLysPheIleIleArgLysSerAnSer 302  
|| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| :: ::  
911 GAAGGTATTATAATGGACTAAAAATTTATTAATAAGAGATATACACCTAAT 960

303 GlnSerIleAsnaspalleValArgLysGluaspTyrlleTyIleuas 319  
::::: ||| :::::: ||| :::::: ||| :::::: ||| :::::: |||  
961 AATCAATA..GATCTTTTGTTAAATCAGGTCATTTTATTAATTA 1005

319 pPhePheAsnLeuasncInGluTrpargValTyThrTyrr..... 332  
::: :::: ||| :::: :::: :::: :::: :::: |||  
0006 .TATGTATCATATAACAATAATGAGCACATGTAGGTTATCCGAAGATG 1054

333 .LysTyPheLysLysGluGluGluLysLeuPheLeuAlaProIleSer 348  
::: ||||| :::: :::: :::: ||| :::: :::: :::  
0005 GAATGCTTTTAATCTTGATAGAACTTACAGTAGGTTATATGCC 1104

349 AspSerAspGluLeuTyAsnThrIleGlnIleLysgluTyrrAspGlugl 365  
::: ||||| :::: :::: :::: :::: :::: :::  
11111105 CAGGTATCCCTCTTTATAAAAAATGGAAGCAGTAAAAATGCGTGATTT 1154

365 nProThyrrSerCysGlnLeuPheLysLysAspGluGluSerThra 382  
||||| ||||| ||||| :::: ||||| :::: :::: :::  
11111155 AAAAAACCATTCTGTCAACACTAAATATATATGATGATATAAAAATGCATCT. 1203

382 spGluIleGlyLeuIleGlyIleHisargPheTyrgluSerGlyIleval 398  
::: ||||| ::||| ||| :::: :::: :::: :::  
1204 .....TTAGAGACTAGTAGGTACCAT.....AATGCTCAAAATAGGC 1239

399 PheGluGluTyrrLysaspTyrrPheCysIleSerLysTyrrTyrrLeuylsgl 415  
::: :::: ||||| :::: ||||| :::: ||||| :::: |||||  
1204 AACGATCCAATAGGGATATATTAAATTTGAAAGCAACGTGGTACTTTTATCA 1289

415 uValLysArgLysProtyrrAsnLeuLysLeuLysCysAsnTyrrGlnPheI 432

```

1290 TTTAAAGATAAA,.....ATTTAGGATGTCATGTGGTACTTTC 1327
      ::::: ||| |||||::: ||| |||
432 leProLysAspGluGlyThr 439
1328 TTCCAACCGATGAAGTTGGACC 1350
      ::::: |||||::: |||

```

OM of: US-09-910-186a-8 to: EST:\* out\_format : pfs

Date: Sep 2, 2002 3:28 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

#### Command line parameters:

```
-MODEL=framet_p2n.model -DEV=xlh  
-O=/cgn2_1/USPTO_spool/US09910186/runat_29082002_134725_15979/app_query.fasta_1.505  
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-DELEXT=7.000 -XGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -NATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500  
-MINLEN=0 -MAXLEN=2000000000 -USER=US09910186 @CGN1_1_2803  
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30  
-NO_XLPXY -WAIT -THREADS=1
```

#### Search information block:

Query: US-09-910-186a-8  
Query length: 440  
Database: EST:\*  
Database sequences: 13736207  
Database length: 1841457050  
Search time (sec): 1596.540000

#### score\_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_est2:BM273933	+	129.50	241.45	0.0003	587	BM273933 PfESToa61e08.y1 Plasm
gb_est2:BM275542	+	128.00	239.11	0.0005	550	BM276542 PfESToa64g04.y1 Plasm
gb_est2:BM275974	+	128.00	238.93	0.0005	559	BM275974 PfESToa68e05.y1 Plasm
gb_est2:BM275232	+	128.00	238.93	0.0005	560	BM275232 PfESToa65h03.y1 Plasm
gb_est2:BM170629	+	123.00	228.87	0.0017	565	BM170629 EST573152 PyBS Plasmod
gb_est2:BM164577	+	123.00	224.81	0.0028	628	BM164577 EST567100 PyBS Plasmod
gb_est1:AL557249	+	114.00	205.54	0.0334	963	AL557249 LTI_FL012_TCL
gb_gss:BM150480	+	113.00	204.15	0.0400	907	BM150480 ENQB95TFB Entamoeba h
gb_est2:BM276398	+	112.50	210.47	0.0178	439	BM276398 PfESToa82f02.y1 Plasm
gb_est1:AL557255	+	112.50	202.69	0.0482	950	AL557255 LTI_FL012_TCL
gb_gss:AS254527	+	112.00	205.64	0.0330	642	AS254527 233pbc08 Pb MBN #21 Pl
gb_est2:BG603391	+	111.50	205.28	0.0346	603	BG603391 EST502481 Plasmodium y
gb_gss:BA543173	-	111.50	203.45	0.0437	723	BA543173 BOKB67TR BOKB Brassic
gb_est2:BM163086	+	111.00	201.28	0.0578	812	BM163086 EST565609 PyBS Plasmod
gb_gss:AS549268	-	110.50	199.89	0.0690	844	AS549268 ENTD208TF Entamoeba h
gb_est2:BM167654	+	109.00	197.25	0.0968	815	BM167654 EST562778 PyBS Plasmod
gb_gss:BM160255	+	108.50	197.66	0.0919	709	BM160255 EST562778 PyBS Plasmod
gb_gss:BM158617	-	107.00	192.54	0.1773	876	BM158617 ENTE16TR Entamoeba h
gb_gss:AS253232	+	107.00	192.15	0.1862	910	AS253232 ENTCG60TF Entamoeba h
gb_gss:AS253372	+	107.00	192.15	0.1862	910	AS253372 ENTCX84TF Entamoeba h
gb_est2:BM164327	+	106.50	192.43	0.1797	802	BM164327 EST566850 PyBS Plasmod
gb_gss:BM159101	+	106.50	191.24	0.2092	902	BM159101 ENTRW22TF Entamoeba h
gb_est2:BM153339	+	106.50	190.89	0.2189	934	BM153339 T3 end of clone BDOAAC
gb_est2:BM168385	+	106.00	196.00	0.1137	510	BM168385 PfESToa61e03.y1 Plasm
gb_gss:CN5061KD	-	106.00	192.26	0.1838	739	BM168385 EST570908 PyBS Plasmod
gb_gss:AS268921	+	106.00	190.05	0.2440	920	AL404291 T7 end of clone AT0AAC
gb_est2:BM168450	+	105.50	189.72	0.2543	950	AL689211 ENTRILTR Entamoeba h
gb_est2:BM159176	+	105.50	191.30	0.2077	736	BM168450 EST570973 PyBS Plasmod
gb_est2:BM169979	+	105.50	191.04	0.2147	755	BM159176 EST561699 PyBS Plasmod
gb_gss:AS2532106	-	105.50	190.99	0.2162	759	BM169979 EST572502 PyBS Plasmod
gb_gss:CN506161	-	105.50	189.59	0.2587	872	AS2532106 ENTCJ29TR Entamoeba h
gb_est1:AW688370	+	105.00	189.41	0.2648	888	AL398887 T7 end of clone AS0AAC
gb_est2:BM162427	+	105.00	191.13	0.2123	678	AL688370 NF006603RIF1000 Devel
gb_gss:AO449733	-	104.50	189.87	0.2496	696	BM162427 EST564950 PyBS Plasmod
gb_gss:BM150734	+	104.50	189.25	0.2702	740	AO449733 500003H03.x2 CpiOWAM13
gb_gss:AS253737	+	104.50	187.55	0.3361	876	BM150734 ENTPF74TF Entamoeba h
gb_gss:CN50638V	+	104.50	187.24	0.3495	903	AS253737 ENTBJS1TR Entamoeba h
gb_htc:AK003217	+	104.50	185.81	0.4201	1041	AL401285 T3 end of clone AS0AAC
gb_est2:BM169075	+	104.00	181.60	0.7206	1580	AK003217 Mus musculus 18 days
			189.23	0.2711	672	BM169075 EST571598 PyBS Plasmod

gb\_gss:BM136854 + 104.00 186.69 0.3752 864 ! BHL36854 ENT0032TR Entamoeba  
gb\_gss:AS2534894 + 104.00 186.57 0.3808 874 ! AS2534894 ENTRU16TR Entamoeba  
gb\_est2:BM274388 + 104.00 172.00 2.47 3707 ! C83838 C83838 Dictyostellium  
gb\_est2:BM273951 + 103.00 189.12 0.2746 557 ! BM274388 PfESToa61e08.y1 Pl  
seq\_name: gb\_est2:BM273933 583 ! BM273951 PfESToa61e08.y1 Pl

#### seq\_documentation\_block:

LOCUS BM273933 587 bp mRNA linear EST 20-DEC-2001  
DEFINITION PfESToa61e08.y1 Plasmodium falci-parum 3D7 gametocyte cDNA library  
Plasmodium falci-parum 3D7 cDNA 5', mRNA sequence.  
ACCESSION BM273933  
VERSION 1 GI:17967234  
KEYWORDS EST.  
SOURCE Plasmodium falci-parum 3D7.  
ORGANISM Plasmodium falci-parum 3D7  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 587)  
AUTHORS Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,  
Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,  
Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,  
Tsagarelashvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A.,  
Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,  
Waterston, R., Wilson, R. and Sibley, D.  
WashU Plasmodium EST Project  
Unpublished (2001)  
Contact: L. David Sibley  
WashU Plasmodium EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu

#### TITLE

#### JOURNAL

#### COMMENT

Library was constructed by R. Haywood. DNA sequencing by:  
Washington University Genome Sequencing Center for information on  
obtaining a clone please contact: L. David Sibley  
(sibley@borcim.wustl.edu), Washington University  
Seq primer: -40UP from Gibco  
High quality sequence stop: 424.  
Location/Qualifiers  
1..587  
/organism="Plasmodium falci-parum 3D7"  
/db\_xref="taxon:36329"  
/clone\_lib="Plasmodium falci-parum 3D7 gametocyte cDNA  
library"  
/dev\_stage="gametocyte (stage III-V)"  
/lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
/note="Vector: pBluescript SK plus; Site 1: EcoRI; Site 2:  
XhoI. The library was constructed by R Haywood. cDNAs were  
synthesized from gametocyte poly(A)+ RNA by oligo d(T)  
priming, size-selected and directionally cloned into the  
EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR  
lambda vector (Stratagene). The primary library was mass  
excised as phagemid using the ExAssist helper phage  
(Stratagene). Clones were mass excised using the ExAssist  
helper phage (Stratagene), the phagemids were precipitated  
with PEG 8000 and extracted with phenol/chloroform.  
Phagemid DNA was electroporated into DH10B cells. Clone  
Availability: David Sibley, Washington University."

#### FEATURES

##### source

BASE COUNT 298 a 56 c 57 g 176 t

##### ORIGIN

alignment\_scores:

Quality: 129.50 Length: 208

Ratio: 1.136 Gaps: 9

Percent Similarity: 54.808 Percent Identity: 24.519

##### alignment\_block:

US-09-910-186A-8 x BM273933 ..

Align seg 1/1 to: BM273933 from: 1 to: 587

Unpublished (2001)  
Contact: L. David Sibley  
WashU Plasmidium EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library was constructed by R. Haywood. DNA sequencing by:  
Washington University Genome Sequencing Center For Information on  
obtaining a clone please contact: L. David Sibley  
(sibley@borcim.wustl.edu), Washington University  
Seq primer: -400P from Gibco  
High quality sequence set: 424.

Location/Qualifiers  
1 550

```
1. .550
   /organism="Plasmodium falciparum 3D7"
   /db_xref="taxon:36329"
   /clone_lib="Plasmodium falciparum 3D7 game
   library"
   /dev_stage="gametocyte (stage III-V)"
```

Availability:	David Sibley,	166 t
a	50 c	54 g

alignment_scores:		
Quality:	128.00	Length: 184
Ratio:	1.255	Gaps: 8
Percent Similarity:	55.435	Percent Identity: 25.543

```

161 AsnAsnAlaLysIleTyrIleAsnGlyLysLeuGluSerAsnThrAspIle 177
||||| ||| |||||||| ||| ::| |||:::|
16 AATAAATAACAAGTGTACATAAAATGTTAAGAACTCATTTAATAATGAGAG 65
177 eLysAspIleArgGluValIleAlaAsnGlyGluIleIlePhe..... 191
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
66 AAATGATATATAAACAAATGCAAAAAAATAGATTATATCACTATTATAAATA 115
192 .....LysLeuAspGlyAspIleAspArgThrGlnPheIleTrpMetLys 206
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
116 AATTAAAAAATAATGGAAAT.....AAG 138
207 TyrPhe..... 208
||||::|
139 TATTATGATACCTGTCAAAATAATATATGTAAGGAAACAACAAGTTGA 188
209 .SerIlePheAsnThrGluLeuSerGlnSerAsnIleGluArgTyrIle 225
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
189 TAATATATATAGAACAAATATGAGTACGCTCTTCCAAAAAATATATGAATA 238
225 ysIleGlnSerTyrSerGluTyrLeu.....LysAsp 235
ATATGAATTCATTGGTAAAGTATACCATAAACAAATATTTATAAAGAAATAAT 288
236 PheTroglyv.....AsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250

```

[illegible]

seq name: qb est2:BM276542

seq_documentation_block:					
LOCUS	BM276542	550 bp	mRNA	linear	EST 20-DEC-2001
DEFINITION	Plasmidom falciparum 3D7 gametocyte cDNA library				
	Plasmidom falciparum 3D7 cDNA 5', mRNA sequence.				

RECESSION  
VERSION  
BM276542.1  
GI:17969903

351.	Plasmodium falciparum 3D7.
SOURCE	Plasmodium falciparum 3D7
ORGANISM	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium
REFERENCE	1 (bases 1 to 550)

AUTHORS	Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Page, D., Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I., Tsagarashvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A., Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R. and Sibley, D.
TITLE	WASHU Plasmodium EST Project

**TITLE**







/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Pyl17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybrizAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybrizAP vector and plasmid DNA isolated."

BASE COUNT 302 a 36 c 64 g 163 t  
ORIGIN

## alignment\_scores:

Quality: 123.00 Length: 220  
Ratio: 1.042 Gaps: 13  
Percent Similarity: 53.636 Percent Identity: 25.000

## alignment\_block:

US-09-910-186A-8 x BM170629 ..

Align seg 1/1 to: BM170629 from: 1 to: 565

```
212 AsnThrGluLeuSerGlnSerAsnIleGluGluArgTyrLys..... 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3 AATAATCAATAAACAACGACGAAAGAGTAATACAAAAAAGA 52

226 .IleGlnSerTyrSerGluTyrLeuLysAspPheTyrGlyAsnProLeuM 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53 AATGAATAGTAGAGGAAATATATAGAAAAATCTTCATGAAGAAATCA 102

242 etyTyrAsn...LysGluTyrTyrMetPheAsnAlaGlyAsnLysAsnSer 257
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 TTATTACACAAAANAAGTATTAGCAATAAATAGACAGGTTTAAAT... 149

258 TyrIleLysLeuLysLysAspSerProValIcylGluIleLeuThrArgSe 274
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 ...TTAAAAAGAAAAAAGAAAAAATGCCCCAAAAAAGAGAGAA 196

274 rLysTyrAsnGlnAsnSerLysTyrIleAsnTyr.....ArgAspLeuT 289
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
197 AAAAAAATTCAAAGTAAAAATCATTTATATTTTAAAAAGACATAT 246

289 yrIleGlyGluLysPheIleIleArgArgLysSer.....AsnSer 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
247 AT.....TTTTAAGAAAAAAGAGAGTTTCATATAAATAA 284

303 GlnSerIleAsnAspIleValArg.....LysGluAs 314
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
285 GATAATTTAAGTAAACATATAGTAAACATAAAACATAACGAGAGAA 334

314 pTyrIleTyrLeuAspPheAsnLeuAsnGlnGluTyrArgValTyrT 331
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
335 ATATGATTTTATTGATAAAGATATTTA..... 362

331 hrTyrLysTyrPheLysGluGluLysLeuPheLeuAlaProIle 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
363 .....TATAAATAAATTTAGAGCATTA..... 386

348 SerAspSerAspGluLeuTyrAsnThrIleGlnIleLysGluTyrAspG 364
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
387 ...AATTCATAGAAAAATATATCATTTTGTAGTGATATATATACAG 433

364 uGlnProThrTyrSerCysGlnLeuLeuPheLysLysAspGluGluSert 381
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
434 TTATAAGGGATATACA.....GTGTTGTGTTT.....GATGACAATTTTC 471
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
381 hrAspGluIleGlyLeuIleGlyIleHisArgPheTyrGluSerGlyIle 397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
472 GAGAT.....TATATTAATCAAAAGTA 494

398 ValPheGluGluTyrLysAspTyrPheCysIleSerLysTyrTyrLeuLy 414
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
495 ATAAATTTAGATATAAAGATTAT.....AAATATTTGATGGA 532

414 sGluValLys 417
|||||:|||||:|||||:|||||:|||||:|||||:
533 TTCATTAAAA 542
```

seq\_name: gb\_est2:BM164577

## seq\_documentation\_block:

LOCUS BM164577 628 bp mRNA linear EST 04-DEC-2001  
DEFINITION EST567100 PyBS Plasmodium yoelii yoelii cDNA clone PYCLW17 5' end,  
mRNA sequence.  
ACCESSION BM164577  
VERSION BM164577.1 GI:17310258  
KEYWORDS EST.  
SOURCE Plasmodium yoelii yoelii.  
ORGANISM Plasmodium yoelii yoelii.  
REFERENCE 1 (bases 1 to 628)  
AUTHORS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,  
Fraser,C.M. and Carucci,D.J.  
Plasmodium yoelii EST project at TIGR  
Unpublished (2001)  
Contact: Jane Carlton  
Parasite Genomics Group  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-530-9319  
Fax: 301-838-0208  
Email: carlton@tigr.org  
For clone info, please contact the Malaria Research and Reference  
Reagent Resource Center, NRC  
http://www.malaria.mr4.org/mr4pages/index.html  
Seq primer: ADF.

## FEATURES

source

1..628  
/organism="Plasmodium yoelii yoelii"  
/strain="17XL"  
/db\_xref="taxon:73239"  
/clone\_lib="pyBS"  
/dev\_stage="Asexual blood stages"  
/lab\_host="E. coli XL-1 Blue"  
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Pyl17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybrizAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybrizAP vector and plasmid DNA isolated."

BASE COUNT 304 a 49 c 53 g 222 t  
ORIGIN

alignment\_scores:

```

Quality: 121.50          Length: 248
Ratio: 0.980             Gaps: 9
Percent Similarity: 50.000 Percent Identity: 23.387

alignment_block:
US-09-910-186A-8 x BM164577

Align seg 1/1 to: BM164577 from: 1 to: 628

6 AsnSerGluLeuLeuAsnAsnIlelleLeuAsnLeuArgTyrLysAspAs 22
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 54
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 nAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluValTyrAspG 39
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 AGCAATACTACATAAATGGGAATGACAAAATAATAGTGAACATACAG 104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
39 lyValGluLeuAsnAspLysAsnGlnPheLeuLysLeuThrSerSerAlaAsn 55
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
105 ATAAGATAAATAATGATCAAAATTTGAAATCATTAACAAGC..... 145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
56 SerLysIleArgValThrGlnAsnGlnAsnIlellePheAsnSerValph 72
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
146 .....TTTAATTTAATTTA 159

72 eLeuAspPheSerValSerPheTrpIleArgIleProLysTyrLysAsnA 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 T.....GAAGAAATG 170

89 spGlyIleGlnAsnTyrIleHis.....AsnGluTyrThrIleLeuAsn 103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
171 ATGGATTGCTAATATATAGATGGGAATGGAACAAACGTCATTA... 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 CysMetLysAsnAsnSerClyTrpLysIleSerIleArgGlyAsnArgIl 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218 .....AT 219

120 eileTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPhePheG 137
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 ATGTGTAAACATATATATATAATAATTTATACATAATAATAATAATTT 269
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
137 lutyAsnIleArgGluAspIleSerGluTyrIleAsnArgTrpPhe... 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
270 TATATCATATATATCTGAATTTCAAAATATAATGACAATTTTATAGC 319
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 .....PheValThrIleThrAsnAsnLeu.....AsnAs 162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
320 TATGAATTTATCAAAATATCAAAACAAGTGCTTATCATATTTTATAATTT 369
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
162 nAlaLysIleTyrIleAsnGlyLysLeuGluSerAsnThrAspIleLysA 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 CTCAAATTTTTCNAAAAATAAAGATTTCAATATATACATGGAATAAG 419
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 spIleArgGluValIleAlaAsnGlyGluIleIlePheLysLeuAspGly 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
420 TAATAAATAATGTATGTAATGACGAACATATATT...TTAGATTTA 466
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 AspIleAspArgThrGlnPheIleTrpMet.....LysTyrPheSe 209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
467 AATTATGAAAAACAATTTTAAATTTTATTGATATAAAAAAATCATAA 516
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
209 rIlePheAsnThrGluLeuSer.....GlnSerAsnIleGluG 222
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
517 TATTATTACACACCTTTAAATAAATCCATAAATAGACCAAAATATAGATA 566
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 luArgTyrLysIleGlnSerTyrSerGluTyrLysLeuLysAspPhe 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
567 ACAATATAACAATTCGTAATAATATTTTATTATACATTTTAAATTAAC 610
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

seq_name: gb_est1:AL557249
seq_documentation_block:
LOCUS
AL557249
```

```
||||| ||| ||| ::||| ::||| |||||
475 ...CTCAATGTGTT...CCTAATGGGATGATGATGAAGAGAAATAC... 516
248 rMetPheAsnAlaGlyAsnLysAsn.....SerTyrIleL 260
||||| ::||| ::||| ::||| ::||| ::|||
517 .....AGGGTCAGGATCAGTCAAGAGATTCGGCAGATTACATCA 556
260 ysLeuTysLysAspSerProValGlyGluLe..... 270
::||| ::||| ::||| ::||| ::|||
557 GGCAACAAAAAGTGACCCCATTCAGAAATTCGGGACTTAGCAGAAATC 606
271 .....LeuThrArgSerLys.....TyrAsnGlnAsnSe 280
||| ||| ||| ||| ||| |||
607 ACCACTCTTGATCGCAGCAAAAGAAATATCATTTGGATATTTGACGAAAA 656
280 rLysTyrIleAsnTyrArgAspLeuTyrIleGlyGluLysPheIleLea 297
||||| ||||| ::||| ::||| ::|||
657 GGACTCGGCAACTATAGA.....GTTTTG 682
297 rgArgLysSerAsnSerGlnSerIleAsnAspIleValArgLysGlu 313
::||| ::||| ::||| ::||| ::|||
683 AACGAGTAGCGAAT.....ATTTCATGAT 708
314 AspTyrIleTyrLeuAspPheAsnLeuAsnGlnGluTrpArgValTy 330
||| ::||| ::||| ::||| ::|||
709 GACTGTGCCTTCTCTGCTGCTTGGGGATGTTTCAAAACCGGAAGATA 758
330 rThr.....TyrLys.....TyrPheLysLysGluGluL 341
||| ||| ::||| ::||| ::|||
759 TAGTGGCGACAAATAATCTACAAACCCAGGCGATCTGCTCCGGATA 808
341 ysLeuPheLeuAlaProIleSerAspSerAspGluLeuTyrAsnThrIle 357
::||| ::||| ::||| ::||| ::|||
809 TGGTGACTTGGGAGCTATCACAAATTTTCGACTGACTTACAATTCGATT 858
358 GlnIleLysGluTyrAspGluGlnProThrTyrSerCys..... 370
||| ||| |||
859 CAAGATAAA.....TGCTGCTCTCTGT 881
371 ....GlnLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
::||| ::||| ::||| ::||| ::|||
882 CCAGAAATACATTTGAAATGGAGAGGAATTACACAGAAGAA...GGAC 928
386 eulLeGlyIleHis 390
||| |||
929 TGGCCTTTTCTCAT 942
seq_name: gb_gss:BH150480
```

```
seq_documentation_block:
LOCUS BH150480 907 bp DNA linear GSS 27-AUG-2001
DEFINITION ENT0957FB Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION BH150480
VERSION BH150480.1 GI:15312077
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 907)
AUTHORS Loftus.B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@igr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
```

```
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 54
High quality sequence stop: 784.
FEATURES
source
1..907
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site_1: Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
```

BASE COUNT 382 a 91 c 149 g 285 t  
ORIGIN

alignment\_scores:  
Quality: 113.00 Length: 349  
Ratio: 0.661 Gaps: 17  
Percent Similarity: 48.997 Percent Identity: 19.484  
alignment\_block:  
US-09-910-186a-8 x BH150480 ..  
Align seg 1/1 to: BH150480 from: 1 to: 907  
29 GlyTyrGlyAlaLysValGluValTyrAspGlyValGluLeuAsn..... 43  
||| ::||| ::||| ::||| ::||| ::|||  
26 GGTTCCTATATCGTACAACAAAATTCGGGATTACATTTTCATCGTTTCAT 75  
44 ....AspLysAsnGlnPheLysLeuThrSerSerAlaAsnSerLys.Ile 58  
||| ::||| ::||| ::||| ::||| ::|||  
76 TATTGCAAAATCCTGTTTAAAACTAATATAAAAGAAAATTCACACTG 125  
59 ArgValThrGlnAsnGlnAsnIlePheAsnSerValPheLeuAspPh 75  
::||| ::||| ::||| ::||| ::||| ::|||  
126 AAATAGCATTAGATAAAAAAATGAGTATATCATTTACTATTATAAATAAT 175  
75 eSerValSerPheTrp.....IleArgIleProLysTyrLysAsnAspG 90  
::||| ::||| ::||| ::||| ::||| ::|||  
176 TCTTGTTGATATATCAGTCCTCTATAAAACCCACGTTTAGTTCATGTT 225  
90 lTleGlnAsnTyrIleHisAsnGluTyrThrIleIleAsnCysMetLys 106  
||| ::||| ::||| ::||| ::||| ::|||  
226 TTATTCAAGAGGAAATCATTTATGAATATGGAACATAACAACAAATTTGAA 275  
107 AsnAsnSerGlyTrpLysIleSerIleArgGlyAsnArgIleIleTrpTh 123  
::||| ::||| ::||| ::||| ::||| ::|||  
276 GGTGATAAT.....GGTGAAACAGAAATTACAATA 304  
123 rLeuIleAspIleAsnGlyLysThrLysSerValPhePheGluTyrAsnI 140  
::||| ::||| ::||| ::||| ::||| ::|||  
305 TTCTGTTGATATAAACAATATGTTTCATTTGATTTCATGATCAACAA 354  
140 leArg.....GluAspIle.....Ser 145  
||| ::||| ::||| ::||| ::||| ::|||  
355 TAGAAAGTGTGAATGTCAAGAGAGATGATTTGATTGTTTCAACAAACAAT 404  
146 GluTyrIleAsnArgTrp.....PhePheValThrIleThrAsnAs 159  
::||| ::||| ::||| ::||| ::||| ::|||  
405 AAATATATTGAAACATGGAGTAATAATTCATCTGTTTACAGGGTCTAATGA 454





BASE COUNT	337 a	30 c	46 g	190 t
ORIGIN				











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2002, 15:01:23 ; Search time 1635.05 Seconds  
(without alignments)  
11069.641 Million cell updates/sec

Title: US-09-910-186A-7

Perfect score: 1341

Sequence: 1 gaattcacgatgcccaacaa.....ggaccgaatagtagaattc 1341

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*\*  
2: em\_esthum:\*\*  
3: em\_estin:\*\*  
4: em\_estmu:\*\*  
5: em\_estov:\*\*  
6: em\_estpl:\*\*  
7: em\_estro:\*\*  
8: em\_htc:\*\*  
9: gb\_est1:\*\*  
10: gb\_est2:\*\*  
11: gb\_hic:\*\*  
12: gb\_gss:\*\*  
13: em\_gss\_hum:\*\*  
14: em\_gss\_inv:\*\*  
15: em\_gss\_pln:\*\*  
16: em\_gss\_vrt:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70.6	5.3	827	12 CNS02156	AL176451 Tetraodon
2	59.4	4.4	394	9 AU060224	AU060224 AU060224
3	59.2	4.4	641	10 BM181884	BM181884 f551b11.y
4	57.4	4.3	849	12 AZ546009	AZ546009 ENTFW53TF
5	54.6	4.1	542	10 B1594734	B1594734 As_tgz_43
6	53.2	4.0	575	10 BM278279	BM278279 As_tgz_55
7	52.4	3.9	500	12 B67199	B67199 CPG0015B Cp
8	52.2	3.9	612	9 BE195101	BE195101 HVSMEH08
9	52	3.9	970	12 CNS03H6V	AL243904 Tetraodon
10	51.8	3.9	681	12 CNS02EOD	AL193990 Tetraodon
11	51.8	3.9	904	12 CNS03HA0	BM244017 Tetraodon
12	51.6	3.8	546	10 BM278263	BM278263 As_tgz_55
13	50.8	3.8	519	10 BM278164	BM278164 As_tgz_54
14	49.6	3.7	615	9 AU061402	AU061402 AU061402
15	49.2	3.7	522	10 BM278558	BM278558 As_tgz_64
16	48.8	3.6	511	10 BM004465	BM004465 TgESTzya8
17	48.8	3.6	905	12 AZ550256	AZ550256 ENTEV58TR

18	48.6	3.6	450	9 AU060996	AU060996 AU060996
19	48.6	3.6	795	12 AZ528485	AZ528485 ENTCM64TF
20	48.6	3.6	823	12 AZ676218	AZ676218 ENTKE36TR
21	48.6	3.6	843	12 BH139532	BH139532 ENTNG88TF
22	48.2	3.6	900	12 AZ549980	AZ549980 ENTDD94TF
23	48	3.6	628	9 AU060230	AU060230 AU060230
24	48	3.6	878	12 BH153470	BH153470 ENTREI9TF
25	47.8	3.6	1016	12 CNS04ESQ	AL287459 Tetraodon
26	47.6	3.5	445	10 BM167469	BM167469 EST569992
27	47.6	3.5	598	10 BM170666	BM170666 EST573189
28	47.6	3.5	605	10 BM163520	BM163520 EST566043
29	47.6	3.5	645	10 BM163530	BM163530 EST567873
30	47.6	3.5	647	10 BM163120	BM163120 EST565643
31	47.6	3.5	671	10 BM160252	BM160252 EST562775
32	47.6	3.5	694	10 BM168242	BM168242 EST570765
33	47.6	3.5	717	10 BM160500	BM160500 EST563023
34	47.6	3.5	747	10 BM162732	BM162732 EST565255
35	47.6	3.5	756	10 BM162492	BM162492 EST565015
36	47.6	3.5	794	10 BM159906	BM159906 EST562429
37	47.2	3.5	477	10 BM278697	BM278697 As_tgz_66
38	47.2	3.5	529	10 BM278587	BM278587 As_tgz_64
39	46.8	3.5	546	10 C22974	C22974 C22974 Dict
40	46.6	3.5	645	9 AI389106	AI389106 GH20192.5
41	46.6	3.5	805	12 AZ753057	AZ753057 RPC1-24-8
42	46.4	3.5	525	10 BM278174	BM278174 As_tgz_54
43	46.4	3.5	890	12 AZ530768	AZ530768 ENTBH54TF
44	46.2	3.4	1101	12 CNS0039G	AL063921 Drosophil
45	46	3.4	548	10 BM159818	BM159818 EST562341

## ALIGNMENTS

### RESULT 1

#### CNS02156

#### LOCUS

#### DEFINITION

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### COMMENT

#### FEATURES

#### source

CNS02156 827 bp DNA linear GSS 12-MAY-2000  
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone  
224F10 of library G from Tetraodon nigroviridis, genomic survey  
sequence.

AL176451 GI:7814508  
GSS: genome survey sequence.  
Tetraodon nigroviridis.  
Tetraodon nigroviridis.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
1 (bases 1 to 827)  
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fzames,C., Fisher,C.,  
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.

Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
Unpublished  
2 (bases 1 to 827)  
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fzames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.

Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
Unpublished  
3 (bases 1 to 827)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetraodon>.

Location/Qualifiers  
1..827  
/organism="Tetraodon nigroviridis"





```

QY 682 aagatccagttctactccgaataacctgaaagacttctgggtaactcgcgtgatgtacaac 741
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 716 CATTCTTACATTTTCATCTAAGTCAAAGGAATCTCTTCGGCATGAACACTTAGTATATA 775
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 742 aagaatactatattgttcaatgctgtgaacaagaactcttaccatcaaaactgaagaagac 801
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 776 TGCACCAACCAATGAATAAAGTCTGTCATGATAACTTTTAAATAAATAAATTAAAGAGAT 835
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 802 tct 804
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 836 TGT 838
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
BI594734 BI594734 542 bp mRNA linear EST 07-SEP-2001
LOCUS As_tgz_43C04_SKPL Ascaris suum adult male testis germinal zone from
DEFINITION Alan Scott Ascaris suum cDNA clone As_tgz_43C04 5', mRNA sequence.
ACCESSION BI594734
VERSION BI594734.1 GI:15498221
KEYWORDS EST.
SOURCE pig roundworm.
ORGANISM Ascaris suum
Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea
; Ascarididae; Ascaris.
REFERENCE 1 (bases 1 to 542)
AUTHORS N., Quayle, M. and Barrell, B.
TITLE Edinburgh University/Sanger Centre Nematode EST Project
JOURNAL Unpublished (2000)
COMMENT Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The library was prepared by Michelle Lizotte-Waniewski for Alan
Scott, Johns Hopkins University Medical School, Baltimore.
Sequencing was performed by the Pathogen Sequencing Unit, Sanger
Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell).
PCR Primers
FORWARD: T3
BACKWARD: T7PL
Plate: 43 row: C column: 04
Seq primer: SKPL
High quality sequence stop: 494.
FEATURES
source
1..542
/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone="As_tgz_43C04"
/clone_lib="Ascaris suum adult male testis germinal zone
from Alan Scott"
/dev_stage="Male"
/note="Vector: Lambda Zap II; Site 1: EcoRI; Site 2: XhoI;
Library was made from dissected testis germinal zone from
adult male Ascaris suum collected from abattoirs.
Constructed by Michelle Lizotte-Waniewski for Alan Scott,
Johns Hopkins University Medical School, Baltimore, MD."
BASE COUNT 204 a 152 c 67 g 119 t
ORIGIN
Query Match 4.1%; Score 54.6; DB 10; Length 542;
Best Local Similarity 48.5%; Pred. NO. 0.0013;
Matches 181; Conservative 0; Mismatches 189; Indels 3; Gaps 1;
QY 278 gtatccagaattacatcccaatgaatacacatcatcaactgcgatgaagaataactctg 337
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 44 GCACACAGTGGTCCAACTGGAATAGCAACAGCAACTTCAATAGCAATTTCAACAACAACA 103
QY 338 gttggaagatctccatcccggttaacggtatcatctggactctgcatcgatatacaacggta 397
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104 ATTTCAACACACCAACTTCAACAACAACACTTCAACAGCAATTTCAACAACAACAAGCACTTCA 163
QY 398 agacaaatctgtattcttcgaatacaacatccggtgaagacatctctgaatacatcaatc 457
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 A---CAGCTTTTCAACAACAACACTTCAACAGCAATTTCAACAACAACAAGCACTTCAACAGCA 220
QY 458 gctggtctctgttaccatcaacaataaactgaacaaatgctaaaaatctacatcaacggta 517
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 221 ATTTCAGCAGCAGCAGCTTCAGCAATTAAGTCAACAACCTTCAACAACCTTCAACAACCTCAGTA 280
QY 518 aactggaatctaataccgagatcaaaagacatccggtgaagttatcgctgaacggtaaatca 577
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 281 ACTTGAAACAATTCACAACAACCTTCAACAACCTTCAACAACCTTCAACAACAACCTCAGTA 340
QY 578 tcttcaaaactggagcgtgacatcgatcgaccagttcatctggtgaataacttctcca 637
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 341 ACCTCAACAACAACCTTAGCAACCTTCAGCAACAGTAACTTCAACAACAACAAGCACTTCAACA 400
QY 638 tcttcaaacacccga 650
Db 401 GCAACTTCAACAA 413

RESULT 6
BI278279 BI278279 575 bp mRNA linear EST 20-DEC-2001
LOCUS As_tgz_55F07_SKPL Ascaris suum adult male testis germinal zone from
DEFINITION Alan Scott Ascaris suum cDNA clone As_tgz_55F07 5', mRNA sequence.
ACCESSION BI278279
VERSION BI278279.1 GI:17971537
KEYWORDS EST.
SOURCE pig roundworm.
ORGANISM Ascaris suum
Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea
; Ascarididae; Ascaris.
REFERENCE 1 (bases 1 to 575)
AUTHORS Blaxter, M.L., Parkinson, J., Whitton, C., Daub, J., Guiliano, D., Hall
N., Quayle, M. and Barrell, B.
TITLE Edinburgh University/Sanger Centre Nematode EST Project
JOURNAL Unpublished (2000)
COMMENT Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The library was prepared by Michelle Lizotte-Waniewski for Alan
Scott, Johns Hopkins University Medical School, Baltimore.
Sequencing was performed by the Pathogen Sequencing Unit, Sanger
Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The
sequence contained a PolyA tail (trimmed)
PCR Primers
FORWARD: T3
BACKWARD: T7PL
Plate: 55 row: F column: 07
Seq primer: SKPL
High quality sequence stop: 494.
FEATURES
Location/Qualifiers
1..575
/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone="As_tgz_55F07"
/clone_lib="Ascaris suum adult male testis germinal zone
from Alan Scott"
/dev_stage="Male"
/note="Vector: Lambda Zap II; Site 1: EcoRI; Site 2: XhoI;

```



Library was made from dissected testis germinal zone from adult male *Ascaris suum* collected from abattoirs.  
 Constructed by Michelle Lizotte-Waniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore, MD."

```

BASE COUNT      204 a 148 c 74 g 149 t
ORIGIN

Query Match      4.0%; Score 53.2; DB 10; Length 575;
Best Local Similarity 48.1%; Pred. No. 0.0032;
Matches 182; Conservative 0; Mismatches 193; Indels 3; Gaps 1;

QY 273 cgagcggtatccagaattacatccacaaatgaatacacaccatcatcaactgcgatgaagataa 332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54 CGTCGGGTCCTCAAAAGACCACCTGGAAATAGCAACAGCAACTTCAATTGTTTTTTCACAA 113

QY 333 ctctggttggaagatctccatccgcggtaaccgtatcatctgactctgctgatatacaa 392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114 CAACAATTTCAACAACAGCACCCTCAACAACAACACTTCAACAGCAATTTCAACAACAGCAA 173

QY 393 cggtaagaccaaatctgtattcttcgaatacaaacatccggtgaagacatctctgaatacat 452
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 CTTCACAGCAACTTCAATTTATCTTTTTTTTAAATGCGCAACAACATTTTTTGGAGAG 233

QY 453 caatcgctggttcttctgttaccatcaccaataacactgaacatgtctaaatctacatcaa 512
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 CACCTTCAGTAGCAGC---AGCTTCAGCAATTACCTCAACAACATATTTTAAACAACCTCAA 290

QY 513 cggtaactggaatctaataccagacatcaagaacatcccgtaaggttatcgctgaacgtga 572
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291 CTATAACTTGAACAATCTCAACAACCTCTAATAACAACCTCAACAACCTCAACAACCAACCT 350

QY 573 aatcatcttcaactgagcgtgaatcgatcgatccacgttcatctctggtgatgaacttt 632
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 351 CAGTAACCTCAACAACACCTTTAGCAACCTCAGACAGCAAGTAACCTTCAACAACAGCAACTT 410

QY 633 ctccattctcaacaccca 650
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 CAACAGCAACTTCAACAA 428

RESULT 7
LOCUS      B67199          500 bp      DNA      linear      GSS 12-MAY-2000
DEFINITION CPG0015B CpiOWAGDNa1 Cryptosporidium parvum genomic, DNA sequence.
ACCESSION  B67199
VERSION     B67199.1 GI:2642691
KEYWORDS   GSS.
SOURCE     Cryptosporidium parvum.
ORGANISM   Cryptosporidium parvum
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Cryptosporidiidae; Cryptosporidium.
REFERENCE  Strong, W.B. and Nelson, R.G.
            Preliminary profile of the Cryptosporidium parvum genome: an
            expressed sequence tag and genome survey sequence analysis
            Mol. Biochem. Parasitol. 107 (1), 1-32 (2000)
20183851
JOURNAL    20183851
MEDLINE
AUTHORS    Contact: Nelson, R. G.
            Depts. of Medicine & Pharmaceutical Chemistry
            San Francisco General Hospital-University of California, San
            Francisco
            Box 0811, San Francisco, CA 94143-0811, USA
            Tel: 415 206 8846
            Fax: 415 206 3353
            Email: malaria@itsa.ucsf.edu
            Submitted sequence has been edited to remove vector sequences 5' to
            the insert, to correct miscalled bases and assign uncalled (N)
            bases throughout the sequence, and to terminate when base-calling
            became ambiguous.
            Seq primer: T3
            Class: shotgun
            High quality sequence stop: 500.

```

# FEATURES

source

Location/Qualifiers

1..500  
 /organism="Cryptosporidium parvum"  
 /strain="IOWA"  
 /db\_xref="taxon:5807"  
 /clone\_lib="CpiOWAGDNa1"  
 /lab\_host="E. coli XL2 Blue MRF"  
 /note="Vector: pBlueScript II (SK-); Site 1: EcoRV; C.  
 parvum (IOWA isolate) genomic DNA was hydrodynamically  
 sheared to produce fragments having a tight size  
 distribution between 2-4 kb by Dr. Ivonne Thorstenson of  
 the Stanford DNA Sequencing and Technology Center  
 (http://sequence-www.stanford.edu/group/techdev/shear.htm  
 ). The randomly sheared gDNA was chromatographed on  
 Sephacryl S-400 to remove any small fragments and DNA  
 eluting in the void volume was subcloned into an EcoR  
 V-digested, alkaline phosphatase-treated pBlueScript II  
 (SK-) vector and transformed into E. coli strain XL2 Blue  
 MRF'. Recombinant clones from the first plating of the  
 library were selected for sequence analysis using T3 and  
 T7 primers."

BASE COUNT 221 a 195 c 12 g 72 t  
 ORIGIN

Query Match 3.9%; Score 52.4; DB 12; Length 500;  
 Best Local Similarity 45.4%; Pred. No. 0.0049;  
 Matches 188; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

```

QY 625 aaatactctccatcttcaacaccgaactgtctcagtcctcaatctcgaagaacggtacaag 684
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83 AACTACAACAACACTACAACAACACTACAACCCTACGACTACCTACTACTACAACAACACTACTAC 142

QY 685 atccagttcttccgaataacctggaagactctctgggttaatccctgctggtgatacaacaaa 744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 TACCACCACTACTACGACACGACACCAACCACTACTACAACAACAACACTACTACTACTAAC 202

QY 745 gaatactatgttcaatgctggtgaagaactcttcatcaactgaagaagaagactct 804
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 CACCCTACTACAACAACAACACTACTACCCTACTACTACAACAACAACAACACTACTACTACTAC 262

QY 805 ccggttggtgaatctcctgactcgttccaaatacaacacagaaactcttaatacatcaactac 864
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 TACTACTACGACAACAACAACACTACTACTACCACAACAACAACAACAACAACACTACTACTAC 322

QY 865 cgcgacctgtacatcgcgtgaaagttcactccgtcgcaaatcttaactctcsgtccatc 924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 TACCCTACTACTACTACACCACTACCACTACTACTACGACACCAACAACAACAACACTACCACCA 382

QY 925 aatgatgacatcgtacgtaaagaagactacatctcactggactcttcaacctgaatcag 984
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 383 TACTACAACCAACAACAACCACTACTACTACAACCACTACTACTACAACCACTACTACTACTAC 442

QY 985 gaatggtggtatcacacctcaagaagtcttcaagaagaagaagaagcttttc 1038
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 TACTACCACAACAACAACAACACTACTACTACTACTACTACGACGACGACCTACCACCTAC 496

```

RESULT 8

LOCUS BE195101

DEFINITION

HVSMeh0088E19f Hordeum vulgare 5-45 DAP spike EST library

HVCNDA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMeh0088E19f,

mRNA sequence.

BE195101

ACCESSION

BE195101

VERSION

BE195101.3 GI:16321082

KEYWORDS

EST.

SOURCE

barley.

ORGANISM

Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticaceae; Hordeum.

REFERENCE

1 (bases 1 to 612)





Qy 322 atgaagaataactctggttggaagatccatccatccgcggttaacogtatactatctggaactctg 381  
 Db 358 MACAACAGTAGTAATAATAGCAACACCAACAGCAACAAACAAACAAACAAACAGCAGCAACAAAC 417  
 Qy 382 atcgatataacggtgaagacaaatctgta--ttcttcgaatacaacatccggtgaagaca 439  
 Db 418 AACAAACAAACAGCAACCAACAAACAGCAACAAACAAACAAACAAACAGCAGCAACAAAC 477  
 Qy 440 tctctgaatacatatcgctggttctctggttaccatcaccaataaacctgaacaaatgcta 499  
 Db 478 ACAAACAGCAACAAACAAACAAACAGCAGCAACAAACAAACAAACAGCAACAAACAAAC 537  
 Qy 500 aaatctacatacaacggttaactggaatctataatccgacatacaagacatccggtgaagtta 559  
 Db 538 ACAGCAACAGCAGCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAGCA 597  
 Qy 560 tcgctaacggtgaatcatcttcaactggaactgacgtgacatcga 602  
 Db 598 ACACAAACAAACAAACAGCAACAAACAAACAAACAAACAGCAACAGCA 640

## RESULT 12

BM278263 546 bp mRNA linear EST 20-DEC-2001  
 LOCUS AS\_tgz\_55D05\_SKPL Ascaris suum adult male testis germinal zone from  
 Alan Scott Ascaris suum cDNA clone As\_tgz\_55D05 5', mRNA sequence.

## ACCESSION

BM278263

## VERSION

BM278263.1 GI:17971521

## KEYWORDS

EST.

## SOURCE

pig roundworm.

## ORGANISM

Ascaris suum  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea  
 ; Ascarididae; Ascaris.

## REFERENCE

1 (bases 1 to 546)

## AUTHORS

Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Guiliano,D., Hall

## TITLE

Edinburgh University/Sanger Centre Nematode EST Project

## JOURNAL

Unpublished (2000)

## COMMENT

Contact: Blaxter ML  
 Institute of Cell, Animal and Population Biology  
 University of Edinburgh  
 Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
 3JT, UK.  
 Tel: +44 131 650 6760  
 Fax: +44 131 670 5450  
 Email: mark.blaxter@ed.ac.uk  
 The library was prepared by Michelle Lizotte-Waniewski for Alan  
 Scott, Johns Hopkins University Medical School, Baltimore.  
 Sequencing was performed by the Pathogen Sequencing Unit, Sanger  
 Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The  
 sequence contained a PolyA tail (trimmed)

## PCR Primers

FORWARD: T3

BACKWARD: T7PL

Plate: 55 row: D column: 05

Seq primer: SKPL

High quality sequence stop: 546.

## FEATURES

source

1..546  
 Location/Qualifiers  
 /organism="Ascaris suum"  
 /db\_xref="taxon:6253"  
 /clone="As\_tgz\_55D05"  
 /clone\_lib="Ascaris suum adult male testis germinal zone  
 from Alan Scott"  
 /sex="Male"  
 /dev\_stage="Adult"  
 /note="Vector: Lambda Zap II; Site\_1: EcoRI; Site\_2: XhoI;  
 Library was made from dissected testis germinal zone from  
 adult male Ascaris suum collected from abattoirs.  
 Constructed by Michelle Lizotte-Waniewski for Alan Scott,  
 Johns Hopkins University Medical School, Baltimore, MD."

## BASE COUNT

206 a 151 c 68 g 121 t

## ORIGIN

Query Match 3.8%; Score 51.6; DB 10; Length 546;  
 Best Local Similarity 48.3%; Pred. No. 0.0083;  
 Matches 175; Conservative 0; Mismatches 184; Indels 3; Gaps 1;

Qy 289 tacatccacaatgaatacacaccatcaatcaactgcatggaataaactctggttggaagatc 348  
 Db 41 TCCAACCTGGAATAGCAACAGCTTTTTCATAGCAATTTCAACAAACACCCCTTTCAACAA 100  
 Qy 349 tccatccgcggttaacogtatactggaactctgatactatcaacggttaagacaatct 408  
 Db 101 AGCAACTTCAACCAACAACTTCAACAGCAATTTCAACAAACAGCAACTTCAACAGCAACT-- 158  
 Qy 409 gtattctcgaatacaacatccggtgaagacatctctgaatacatatcaatcggtgttcttc 468  
 Db 159 -TCAACAAACAACTTCAACAGCAATTTCAACAAACAGCAACTTCAACAGCAATTTTCAGCAGC 217  
 Qy 469 gttaccatcaccaataacctggaacaaatgctaaaatctacatacaacggttaacgtgaatct 528  
 Db 218 ATCAGCTTTCAGCAATTTACCTTCAACAACTCTCTTAAACAACTCAACTATAACTTTGAACAAT 277  
 Qy 529 aataccgacatacaagacatccggtgaagtattcgtcaacggtgaatacatcttcaaaactg 588  
 Db 278 CTCAACAACTCTCAATAACAACCTCAACAATCTCAACAAACAACTCACTAACCCTCAACAAAC 337  
 Qy 589 gacggtgacatcgatcgatcccggttctcatctggtgatgaatactctccatcttcaacacc 648  
 Db 338 AACCTTAGCAACCTCAGCAACAGTAGTAACCTTCAACAAACAGCAACTTCAACAGCAACTTCAAC 397  
 Qy 649 ga 650  
 Db 398 AA 399

## RESULT 13

BM278164

## LOCUS

BM278164 519 bp mRNA linear EST 20-DEC-2001  
 AS\_tgz\_54A11\_SKPL Ascaris suum adult male testis germinal zone from  
 Alan Scott Ascaris suum cDNA clone As\_tgz\_54A11 5', mRNA sequence.

## ACCESSION

BM278164

## VERSION

BM278164.1 GI:17971422

## KEYWORDS

EST.

## SOURCE

pig roundworm.

## ORGANISM

Ascaris suum  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea  
 ; Ascarididae; Ascaris.

## REFERENCE

1 (bases 1 to 519)

## AUTHORS

Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Guiliano,D., Hall  
 ,N., Quayle,M. and Barrell,B.

## TITLE

Edinburgh University/Sanger Centre Nematode EST Project

## JOURNAL

Unpublished (2000)

## COMMENT

Contact: Blaxter ML  
 Institute of Cell, Animal and Population Biology  
 University of Edinburgh  
 Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
 3JT, UK.  
 Tel: +44 131 650 6760  
 Fax: +44 131 670 5450  
 Email: mark.blaxter@ed.ac.uk

The library was prepared by Michelle Lizotte-Waniewski for Alan  
 Scott, Johns Hopkins University Medical School, Baltimore.

Sequencing was performed by the Pathogen Sequencing Unit, Sanger  
 Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The  
 sequence contained a PolyA tail (trimmed)

PCR Primers

FORWARD: T3

BACKWARD: T7PL

Plate: 54 row: A column: 11

Seq primer: SKPL

High quality sequence stop: 519.

## FEATURES

source

Location/Qualifiers  
 1..519

```
/organism="Ascaris suum"  
/db_xref="taxon:6253"  
/clone="AS_tgz_54All"  
/clone_lib="Ascaris suum adult male testis germinal zone  
from Alan Scott"  
/sex="Male"  
/dev_stage="Adult"  
/note="vector: Lambda zap II; Site.1: EcoRI; Site.2: XhoI;  
Library was made from dissected testis germinal zone from  
adult male Ascaris suum collected from abattoirs.  
Constructed by Michelle Lizotte-Waniewski for Alan Scott,  
Johns Hopkins University Medical School, Baltimore, MD."  
BASE COUNT      192 a 142 c 65 g 120 t  
ORIGIN
```

```
Query Match      3.8%; Score 50.8; DB 10; Length 519;  
Best Local Similarity 48.6%; Pred. NO. 0.013;  
Matches 170; Conservative 0; Mismatches 177; Indels 3; Gaps 1;  
  
QY 195 gaatcagaaatcattcaactcgcgtattctgacttctgttcttctgattcg 254  
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 64 GAATAGCAACAGCAACTTCAACAGCGATTTCACACAAACAATTTCAACAGCAACTT 123  
  
QY 255 tatccgaaatacagaacgagcggtatccagaattacatccacaatgaatacaccatcat 314  
      || || || || || || || || || || || || || || || || || || || || ||  
Db 124 CAACAACAACATTTCAACAGCAACTTCAACAGCAACTTCAACACAATTTCAACAGCATTC 183  
  
QY 315 caactgcatgaagaataactctggttggaagatctccatccgcggtgaacggtatctctg 374  
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 184 CAACCTTTATCAACTTCAACAGCAATTTAGCAGCAACTTTCAGCAATTTACCTCAACAA 243  
  
QY 375 gactctgacgtatcaacggttaa---gacaaactctgtattctcgaatacaacatccg 431  
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 244 CTCTCTTACACCTCACTATATTAATTTGAACAATCTCAACAACCTCAATACACCTCAA 303  
  
QY 432 tgaagacatctctgaatacatacgtcgtgttcttctgttaccatcaccaataaactgaa 491  
      || || || || || || || || || || || || || || || || || || || || ||  
Db 304 CAATCTCAACAACACCTCAGTAACTCAACAACAACCTTAGCACTTAGCACTAGCACTAA 363  
  
QY 492 caatcgtaaaatctacatacaacggttaactggaatcttaataccgacatca 541  
      || || || || |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 364 CTTCAACAACAGCAACTTCAACAGCAATTTCAACAATTTTACGAATAGCA 413
```

```
RESULT 14  
LOCUS AU061402 615 bp mRNA linear EST 20-MAY-1999  
DEFINITION AU061402 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium  
discoideum cDNA clone SLE112, mRNA sequence.  
ACCESSION AU061402  
VERSION AU061402.1 GI:4882506  
KEYWORDS EST.  
SOURCE Dictyostelium discoideum.  
ORGANISM Dictyostelium discoideum  
REFERENCE 1 (bases 1 to 615)  
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,  
Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,  
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.  
TITLE Developmental cDNA in Dictyostelium discoideum  
JOURNAL Unpublished (1998)  
COMMENT Contact: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan  
Email: d402hu@sakura.cc.tsukuba.ac.jp  
PROJECT = Dictyostelium discoideum cDNA project in Japan.  
FEATURES  
Source 1..615  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db_xref="taxon:44689"
```

```
/clone="SLE112"  
/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"  
/dev_stage="slug"  
BASE COUNT      290 a 176 c 31 g 113 t 5 others  
ORIGIN  
  
Query Match      3.7%; Score 49.6; DB 9; Length 615;  
Best Local Similarity 47.0%; Pred. NO. 0.029;  
Matches 154; Conservative 0; Mismatches 174; Indels 0; Gaps 0;  
  
QY 256 atccggaatacagaacgagcggtattccagaattattcaccataaataacacatccatc 315  
      || || || || |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 262 ACCACCAACAACACACAGCAGCAACCAACCAACATCAACAATGTCCCCACATCATC 321  
  
QY 316 aactgcatgaagaataactctggttgggaagatctccatccgcggttaacggtatctctg 375  
      || || || || || || || || || || || || || || || || || || || || ||  
Db 322 ATATCAACAACAACAATATGCCCCACCATCATCAACAACAACAACAATGCTCATCACC 381  
  
QY 376 actctgacgtatcaacggttaagaccaaatctgtattcttcgaatacaacatccgtgaa 435  
      || || || || |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 382 AACACACAACAACAACAACAACAACAATGTTCACGACATCACCACAACAACAATGTCTCAC 441  
  
QY 436 gacatctctgaatacatcaatcgtcgttcttctgttaccatcaccaataaactgaaat 495  
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 442 CACATCAACAACAACAATCCACAATGTGCGCCACATCACCACAACAATGCTCATCACC 501  
  
QY 496 gctaaaatctacatacaacggttaactggaatcttaataccgacatcaaaagacatccgtgaa 555  
      || || || || |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 502 ATCAACAACAACAATCAACAACAACAACAACACCAATGCAACAATGCAACAACAATCAAC 561  
  
QY 556 gttatcgtcaacggtgaaatcatcttca 583  
      || || || || |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 562 AAATGCTACAACCATCAAAAGTTATTTA 589  
  
RESULT 15  
LOCUS BM278558 522 bp mRNA linear EST 20-DEC-2001  
DEFINITION AS_tgz_64B11_SKPL Ascaris suum adult male testis germinal zone from  
Alan Scott Ascaris suum cDNA clone AS_tgz_64B11 5', mRNA sequence.  
ACCESSION BM278558  
VERSION BM278558.1 GI:17971816  
KEYWORDS EST.  
SOURCE pig roundworm.  
ORGANISM Ascaris suum  
Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea  
REFERENCE 1 (bases 1 to 522)  
AUTHORS Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Guiliano,D., Hall  
N., Quayle,M. and Barrell,B.  
TITLE Edinburgh University/Sanger Centre Nematode EST Project  
JOURNAL Unpublished (2000)  
COMMENT Contact: Blaxter ML  
Institute of Cell, Animal and Population Biology  
University of Edinburgh  
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
3JT, UK.  
Tel: +44 131 650 6760  
Fax: +44 131 670 5450  
Email: mark.blaxter@ed.ac.uk  
The library was prepared by Michelle Lizotte-Waniewski for Alan  
Scott, Johns Hopkins University Medical School, Baltimore.  
Sequencing was performed by the Pathogen Sequencing Unit, Sanger  
Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The  
sequence contained a polyA tail (trimmed)  
PCR Primers  
FORWARD: T3  
BACKWARD: T7PL  
Plate: 64 row: B column: 11  
Seq primer: SKPL  
High quality sequence stop: 522.  
FEATURES  
Location/Qualifiers
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2002, 16:01:38 ; Search time 199.71 Seconds  
(without alignments)  
11528.628 Million cell updates/sec

Title: US-09-910-186a-7  
Perfect score: 1341  
Sequence: 1 gaattcacgatgccacaa.....ggaccgaatagtaagaattc 1341

oring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1341	100.0	1341	21	AAZ87213
2	1341	100.0	1341	22	AAA54485
3	1271.4	94.8	1347	22	AAA54589
4	741.2	55.3	1547	19	AAV30581
5	739.6	55.2	3876	19	AAV30579
6	662.8	49.4	1526	19	AAV30580
7	522.4	39.0	3509	19	AAV26289
8	426.8	31.8	1368	22	AAA54491
9	418.6	31.2	1347	21	AAZ87217

10	263	19.6	1323	22	AAA54483	Botulism toxin hea
11	263	19.6	1326	22	AAA54484	Botulism toxin hea
12	263	19.6	1330	17	AAT29245	Type A neurotoxin
13	263	19.6	1330	19	AAV30571	Clostridium botuli
14	263	19.6	1332	22	AAA54482	Botulism toxin hea
15	263	19.6	1338	21	AAZ87212	DNA encoding synth
16	263	19.6	1351	19	AAV30576	Clostridium botuli
17	263	19.6	1402	17	AAT29246	Type A neurotoxin
18	263	19.6	1402	19	AAV30572	Clostridium botuli
19	242.6	18.1	1313	18	AAT48101	Immunogenic type F
20	242.6	18.1	1314	22	AAA54499	Botulism toxin hea
21	242.6	18.1	1317	22	AAA54490	Botulism toxin hea
22	241.6	18.0	1317	21	AAZ87216	DNA encoding synth
23	227.2	16.9	1332	22	AAA54588	Sequence encoding
24	221.6	16.5	702	21	AAZ87221	DNA encoding BONTA
25	218.6	16.3	1535	19	AAV30596	Clostridium botuli
26	182.2	13.6	1374	22	AAA54487	Botulism toxin hea
27	160	11.9	1278	22	AAA54489	Botulism toxin hea
28	160	11.9	1400	21	AAZ87215	DNA encoding synth
29	160	11.9	1400	22	AAA54488	Botulism toxin hea
30	134.2	10.0	3712	16	AAQ97490	Plasmid pHTAL. N
31	131.4	9.8	1359	12	AAQ12121	Synthetic tetanus
32	129.4	9.6	4366	16	AAQ85424	Plasmid pTECH3-P28
33	129.2	9.6	3754	15	AAQ57880	Intermediate plasm
34	129.2	9.6	3754	16	AAQ85420	Plasmid pTECH1. S
35	129.2	9.6	3769	15	AAQ57881	Intermediate plasm
36	129.2	9.6	3769	16	AAQ97492	Plasmid pTECH2. N
37	129.2	9.6	4377	16	AAQ85421	Plasmid pTECH1-P28
38	123	9.2	1317	21	AAZ87220	DNA encoding nativ
39	123	9.2	1546	19	AAV30575	Clostridium botuli
40	123	9.2	2532	21	AAZ87218	DNA encoding nativ
41	123	9.2	3891	17	AAT29244	C. botulinum type
42	123	9.2	4835	21	AAC64582	BONT/A neurotoxin
43	120.4	9.0	1371	21	AAZ87214	DNA encoding synth
44	120.4	9.0	1371	22	AAA54486	Botulism toxin hea
45	103.2	7.7	1766	16	AAQ97491	Plasmid pTECH2 + h

#### ALIGNMENTS

RESULT 1  
AAZ87213  
ID AAZ87213 standard; DNA; 1341 Bp.  
XX AC AAZ87213;  
XX DT 08-MAY-2000 (first entry)  
XX DE DNA encoding synthetic BoNT serotype B (BONTB) Hc fragment.  
XX KW Botulinum neurotoxin; heavy chain; BONT; serotype B;  
XX KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;  
XX KW VEE; botulism; vaccine; diagnosis; drug screening; ds.  
XX OS Clostridium botulinum.  
XX OS Synthetic.  
XX FH Key Location/Qualifiers  
XX CDS 10..1333  
XX FT /\*tag=a  
XX FT /product="Synthetic botulinum neurotoxin serotype B  
(BONTB) heavy chain C-terminal fragment (Hc)"  
XX PN WO200002524-A2.  
XX PD 20-JAN-2000.  
XX PE 09-JUL-1999; 99WO-US15570.  
XX PR 10-JUL-1998; 98US-0092416.  
XX PR 12-MAY-1999; 99US-0133870.







QY 1261 cgcacacccgtacacactgaaactgggtggtgcaattggcagttcccccgaagacgaaggt 1320  
 Db 1261 cgcacacccgtacacactgaaactgggtggtgcaattggcagttcccccgaagacgaaggt 1320  
 QY 1321 tggaccgaatagtaagaattc 1341  
 Db 1321 tggaccgaatagtaagaattc 1341

## RESULT 3

AAA54589

ID AAA54589 standard; DNA; 1347 BP.

XX

AC AAA54589;

XX

DT 11-APR-2001 (first entry)

XX

DE Sequence encoding botulinum toxin C fragment (serotype B).

XX

KW Botulinum toxin; neurotoxin; heavy chain; recombinant expression;  
 KW recombinant vector; antigen; immune response; vaccine; bacterium;  
 KW infection; ds.

XX

OS Synthetic.

OS Clostridium botulinum.

XX

PN WO200067700-A2.

XX

PD 16-NOV-2000.

XX

PF 12-MAY-2000; 2000WO-US12890.

XX

PR 12-MAY-1999; 99US-0133865.

PR

PR 12-MAY-1999; 99US-0133866.

PR

PR 12-MAY-1999; 99US-0133867.

PR

PR 12-MAY-1999; 99US-0133868.

PR

PR 12-MAY-1999; 99US-0133869.

PR

PR 12-MAY-1999; 99US-0133873.

PR

PR 29-JUL-1999; 99US-0146192.

XX

PA (USSA ) US ARMY MEDICAL RES &amp; MATERIAL COMMAND.

XX

PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;

XX

DR WPI; 2001-016048/02.

XX

PT New nucleic acids encoding the carboxy- or amino-terminal portions of  
 PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
 PT vaccine against botulinum

XX

PS Example 8; Page 38-39; 73pp; English.

XX

CC Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
 CC chain and then posttranslationally nicked, forming a dichain which  
 CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
 CC remain linked by a disulfide bond. Nucleic acids encoding the  
 CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
 CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
 CC expression vectors and expressed in transformed cells to produce  
 CC peptide antigens useful for eliciting an immune response to give  
 CC protective immunity against botulinum neurotoxin, which causes  
 CC botulism. The nucleic acids are expressible in a recombinant  
 CC organism such as *Escherichia coli* or *Pichia pastoris*. The use  
 CC of recombinant nucleic acids are advantageous since it eliminates  
 CC the need to culture large quantities of hazardous toxin-producing  
 CC bacterium. Production yield from the genetically engineered product  
 CC is also high and cost of production is lower. The nucleic acids can  
 CC be derived from Clostridium botulinum serotypes A-G.

XX

SQ Sequence 1347 BP; 439 A; 337 C; 238 G; 333 T; 0 other;

XX

Query Match 94.8%; Score 1271.4; DB 22; Length 1347;

Best Local Similarity 99.6%; Pred. No. 0;  
 Matches 1316; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

QY 15 caacaaatacaattccgaataatcctgaacaaataatcctgaacatcgctgttacaagaacaa 74  
 Db 9 caacaaatacaattccgaataatcctgaacaaataatcctgaacatcgctgttacaagaacaa 68

QY 75 caatctgacgactctgtctgtgttacgggtgctaaagtgtgaagtatacgcacggtgttgaaact 134  
 Db 69 caatctgacgactctgtctgtgttacgggtgctaaagtgtgaagtatacgcacggtgttgaaact 128

QY 135 gaatgacaagaacaccagttccaaactgacctctccgctaactctaagatcgcgtgttactca 194  
 Db 129 gaatgacaagaacaccagttccaaactgacctctccgctaactctaagatcgcgtgttactca 188

QY 195 gaatcagaacatcatcttcaactccgattctcctggaacttctgttctctctggtattcgcg 254  
 Db 189 gaatcagaacatcatcttcaactccgattctcctggaacttctgttctctctggtattcgcg 248

QY 255 tatccgaaatacaagaacgcggtatccagaattacatccacaatgaatacaccatcat 314  
 Db 249 tatccgaaatacaagaacgcggtatccagaattacatccacaatgaatacaccatcat 308

QY 315 caactgcatgaagaataaactctgtgttggaagatctccatcccggtgaaccgttatcatctg 374  
 Db 309 caactgcatgaagaataaactctgtgttggaagatctccatcccggtgaaccgttatcatctg 368

QY 375 gactctgacgatatacaacggttaagacaaatctgtattcttcgaatacacaacatccgtga 434  
 Db 369 gactctgacgatatacaacggttaagacaaatctgtattcttcgaatacacaacatccgtga 428

QY 435 agacatctctgaatacatcaatcgctggtttcttaccatccacaataaactgaacaa 494  
 Db 429 agacatctctgaatacatcaatcgctggtttcttaccatccacaataaactgaacaa 488

QY 495 tgcataaatctacatacaacggttaaaactggaatcttaataccgcacatacaagaacatccgtga 554  
 Db 489 tgcataaatctacatacaacggttaaaactggaatcttaataccgcacatacaagaacatccgtga 548

QY 555 agttatcgtaacggtgaatacatcttcaactggaacggtgaacatcgatcgaccagtt 614  
 Db 549 agttatcgtaacggtgaatacatcttcaactggaacggtgaacatcgatcgaccagtt 608

QY 615 catctggatgaataacttctccatcttcaacacgcgaactgtctcagtcacaataatcgaaga 674  
 Db 609 catctggatgaataacttctccatcttcaacacgcgaactgtctcagtcacaataatcgaaga 668

QY 675 acggtacaaagatccagttcttactccgaataacatccatgaaagacttctgggttaactccgctgat 734  
 Db 669 acggtacaaagatccagttcttactccgaataacatccatgaaagacttctgggttaactccgctgat 728

QY 735 gtacacaaagaataactatgttcaatgctggtgaacgaactcttacatacaactgaa 794  
 Db 729 gtacacaaagaataactatgttcaatgctggtgaacgaactcttacatacaactgaa 788

QY 795 gaaagactctccggtgtgtgaaactcctgactgtgttccaaatacaacaccagaactctaaata 854  
 Db 789 gaaagactctccggtgtgtgaaactcctgactgtgttccaaatacaacaccagaactctaaata 847

QY 855 catcaactaccgcgaactgtacatcggtgaaagttcatccatcccgtaacttaactc 914  
 Db 848 catcaactaccgcgaactgtacatcggtgaaagttcatccatcccgtaacttaactc 907

QY 915 tcagttccatcaatgatgacatcgtaacgtaaaagaagactacatctaccctggactttctcaa 974  
 Db 908 tcagttccatcaatga-gacatcgtaacgtaaaagaagactacatctaccctggactttctcaa 966

QY 975 cctgaatcagaagaatggcgtgtatatacactcaagaacttcaagaagaagaagaagaact 1034  
 Db 967 cctgaatca-gaattggcgtgtatatacactcaagaacttcaagaagaagaagaagaact 1025

QY 1035 ttctcgtgactcgtatctgtattccgcgaactctacacacatccagatcaagaata 1094  
 Db 1035 ttctcgtgactcgtatctgtattccgcgaactctacacacatccagatcaagaata 1094

Db 1026 ttctcgtgctcgatctctgattccgacgaactctacaacaccatccagatcaagaata 1085  
QY 1095 cgacgaacgagcactactcttgccagctgctgttcaagaagaagatgaagaatctactga 1154  
Db 1086 cgacgaacgagcactactcttgccagctgctgttcaagaagaagatgaagaatctactga 1145  
QY 1155 cgaaatcgtgctgattccacgcgtttctacgaactcgtgtatcgtattcgaagaata 1214  
Db 1146 cgaaatcgtgctgattccacgcgtttctacgaactcgtgtatcgtattcgaagaata 1205  
QY 1215 caaagactactctgcatctcaaaatggtacctggaaggagttaaacgcaaacgcgtacaa 1274  
Db 1206 caaagact-cttctgcatctcaaaatggtacctggaaggagttaaacgcaaacgcgtacaa 1264  
QY 1275 cctgaaactgggttgcgaattggcagttccatccgaaagacgaaggttgaccgaatagta 1334  
Db 1265 cctgaaactgggttgcgaattggcagttccatccgaaagacgaaggttgaccgaatagta 1324  
QY 1335 a 1335  
Db 1325 a 1325  
RESULT 4  
AAV30581  
ID AAV30581 standard; DNA; 1547 BP.  
XX  
AC AAV30581;  
XX  
DT 07-DEC-1998 (first entry)  
DE  
DE Clostridium botulinum toxin B fragment C gene in pETHIsb.  
XX Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;  
KW botulism; BotB; ds.  
XX  
OS Clostridium botulinum serotype B Danish strain.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT CDS 108..1526  
FT /\*tag= a  
XX  
PN WO9808540-Al.  
XX  
XX 05-MAR-1998.  
XX  
PF 28-AUG-1997; 97WO-US15394.  
XX  
PR 28-AUG-1996; 96US-0704159.  
XX  
PA (OPHI-) OPHIDIAN PHARM INC.  
XX  
PI Thalley BS, Williams JA;  
XX  
XX WPI: 1998-230234/20.  
DR P-PSDB; AAW68394.  
XX  
XX Host cell containing recombinant expression vector encoding  
PT Clostridium botulinum type B or E toxin - useful to treat humans  
PT and other animals at risk of intoxication with clostridial toxin  
XX  
PS Example 35; Page 303-305; 428pp; English.  
XX  
CC This is the DNA sequence of the Clostridium botulinum serotype B  
CC (Danish strain) toxin fragment C gene contained in plasmid pETHIsb.  
CC The encoded fragment C polypeptide (see AAW68394) has a His-tagged  
CC N-terminal extension. The vector was used to express native  
CC (i.e. non-fusion) soluble C fragment in Escherichia coli host  
CC cells. The invention relates to recombinant proteins derived from  
CC C. botulinum toxins. Methods are provided which allow for the  
CC isolation of soluble recombinant proteins free of significant  
CC endotoxin contamination. Preferred hosts for production of

CC recombinant proteins are E. coli, insect cells and yeast cells.  
CC The recombinant toxins are used as immunogens for the production  
CC of vaccines and antitoxins that are useful in the treatment of  
CC humans and animals at risk of intoxication with clostridial toxin.  
XX  
SQ Sequence 1547 BP; 634 A; 148 C; 263 G; 502 T; 0 other;  
Query Match 55.3%; Score 741.2; DB 19; Length 1547;  
Best Local Similarity 72.5%; Pred. No. 9.9e-195;  
Matches 959; Conservative 0; Mismatches 363; Indels 0; Gaps 0;  
QY 10 atggcccaacaatacacaattccgaatacctgaacaataatcatcctgaacctgcgttacaaa 69  
Db 204 atgtttaataataataatagcgaataatttaataataattattcttaatttaagataaag 263  
QY 70 gacaacaatctgacgtatcgtctgttgaactgctgctaaagtgtgaagtatacgcggtgtt 129  
Db 264 gataataatttaataatgatttatcaggatattggcgcaaggtaggtatgatggagtc 323  
QY 130 gaactgaatgacaagaaccagttcaaacgactccttccgtaactcctgaatcccggttt 189  
Db 324 gagcttaatgataaaaatcaatttaactagttcagcaaatagtagtagtg 383  
QY 190 actcagaatcagaacatcatcttcaactcgtattctcgtgacttctctgtttctctctgg 249  
Db 384 actcaaatcagaataatcatatttaataatgctgtgtccttgatttagcgttaagctttgg 443  
QY 250 attcgtatcccgaaatcacagaacgcggtatccagaattatcaccaataacacacc 309  
Db 444 ataaagaatacctaaataagaatggtgtatacaaaaattatttcaataatgaataaca 503  
QY 310 atcatcaactgacgaagaataaactcgtgttggaagatctccatccgcggtgaaccgtatc 369  
Db 504 ataatattgtatgaaaaataaattccggtcggtgaaataatctatttaggggtaatagta 563  
QY 370 atctgactctgactgatatcaacggtgaagaccacaataatctgtattcttcgaataacaac 429  
Db 564 atatggactttaattgatataaaatggaaaaacccaataatcggatttttttgaataaaca 623  
QY 430 cgtgaagacatctctgaatacatcaatcgtgtgttctcgttaccatcccaataacctg 489  
Db 624 agagaagatataatcagatataaaatagatggttttttgtaactattactaataattg 683  
QY 490 acaaatgctaaaatctacatcaacgcggtaaactggaatctaaataaccgacatcaagaacatc 549  
Db 684 aataacgctaaaatttattatgtaagctgagaatacaataacagatataaagatata 743  
QY 550 cgtgaagtattcgttaacggtgaataatcactctcaaaactgacggtgacatcgatctacc 609  
Db 744 agagaagttattgctaatggtgaataatatttaatttaatttagtgggtgatagataaga 803  
QY 610 cagttcatctggtgaataactctctccatctcaaacaccgaactgctcagtcgaataatc 669  
Db 804 caatttatttggatgaataattttcagttatttttaacggaattaaagtcaatcaaatatt 863  
QY 670 gaagaacggtacaagatccagttcttactcgaataactgaaagacttctggggttaaccg 729  
Db 864 gaagaagatataaaatccaatcattatagcaataatttaaaagatttttggggaatccct 923  
QY 730 ctgagtacaacaagaataactatattgttcaatgctggtgaacaagaactctacatcaa 789  
Db 924 ttaattgacaataaagaataattatactgttttaattgcggggaataaaattcattataaa 983  
QY 790 ctgaagaagaactctccggttggtaaaactcgtgactcgttcccaatacaacaccgaactct 849  
Db 984 ctgaagaagaactcactcgttagtgaaatttttaacagctagcaataataatcaaatctct 1043  
QY 850 aaatacatcaactaccgcgcactgtacatcgttgaaaaagtctcatcgcgtcgaactct 909  
Db 1044 aaataataaattatagagatttatatttggagaaaaattattataagaagaagtca 1103  
QY 910 aactctcagtcctcaatgatgacatcgtagcgtaaagaagactacatctaccctggacttc 969





```
Db 924 ttaagtataataaagaattattatgtttaatgcgggaataaaaaattcatattaaa 983
QY 790 ctgaagaagaactccgtgttggaatccctgactcgttccaaatacaaccagaactct 849
Db 984 cagtgaaagattcatctgtagtgaaattataacacgtagcaaatataatcagaattcc 1043
QY 850 aaatacatcaactaccgcgacgtgtacatcggtgaaagtttcatatccgtcgcaaatct 909
Db 1044 aattataaattataagaattattatattggaagaaaaattattataagaagagagtc 1103
QY 910 aactctcagtcacatgaatgatgacatcgtaacgttaagaagatacatctacctggacttc 969
Db 1104 aattctcaatctataaatgatgatagttagtagaaagaagattatatacatctagatttg 1163
QY 970 ttcaacctgaatcaggaaatggcgtgtatatacacctacaagtaacttcaagaagaagaa 1029
Db 1164 gtacttcaccatgaagagtggaagagtataatgctataataattttaagggaacagaa 1223
QY 1030 aagcttttctcgtccgcatctctgattccgcagcaactctacaacaccatccagatcaa 1089
Db 1224 aaatgttttattattataagtgattcttaataatgaattttataagactatagaataaaa 1283
QY 1090 gaatacagacagacgcgacctactctgtccagctgctgttccaagaagaagatgaatct 1149
Db 1284 gaatatgatgaacagccatcatatagttgtcagttgcttttttaaaaaagatgaagaagt 1343
QY 1150 actgacgaataatcggtcgtgacgttaccacggtttctacgaatctggtatctattcgaa 1209
Db 1344 actgatgataggttggtgattcttccttcacgaatctggtgattttacgtaaa 1403
QY 1210 gaatacaagactactctgcctctccaaatggtacctgaaggaagttaaacgcgaacgc 1269
Db 1404 aagtataaagattttttgtataagtaaatggtacttaaaagaggttaaaaggaaccca 1463
QY 1270 tacaacctgaactgggttgcgaattggcagttccatcccgaaagacgaaggttggaccgaa 1329
Db 1464 tataagtcagaattgggtgattgtaattggcagtttattctctaaagatgaaggttggactgaa 1523
QY 1330 ta 1331
Db 1524 ta 1525
```

## RESULT 7

AAV26289  
ID AAV26289 standard; DNA; 3509 BP.

XX AC AAV26289;

XX 27-JUL-1998 (first entry)

XX Recombinant botulinum neurotoxin type B LH728/B encoding DNA.

XX Botulinum; recombinant; Clostridium botulinum; neurotoxin;

KW immunogen; detection; tetanus; non-toxic; toxin; ds.

XX Synthetic.

OS Clostridium botulinum.

XX FH Key Location/Qualifiers  
CDS 1..3509

FT /\*tag= a  
FT /product= "LH728/B"  
FT /note= "no stop codon given"

XX WO9807864-A1.

PN 26-FEB-1998.

XX 22-AUG-1997; 97WO-GB02273.

XX 13-DEC-1996; 96GB-0025996.

PR 23-AUG-1996; 96GB-0017671.  
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
PA (SPEY-) SPEYWOOD LAB LTD.

XX PI Foster KA, Quinn CP, Shone CC;  
XX P-PSDB; AAW56017.

XX WPI; 1998-169168/15.  
XX P-PSDB; AAW56017.

XX Recombinant neurotoxin polypeptides - used to develop therapeutic  
PT agents, immunogens or as non-toxic standards for the detection of  
PT neurotoxins

XX Disclosure; Page 87-91; 137pp; English.

XX The present sequence encodes a recombinant neurotoxin protein from  
CC the present invention. The present invention describes recombinant  
CC neurotoxin proteins which comprise a first and second domain, where  
CC the first domain is adapted to cleave one or more vesicle or  
CC plasma-membrane associated proteins essential to exocytosis, and where  
CC the second domain is adapted: (a) to translocate the protein into a  
CC cell; (b) to increase the solubility of the protein compared to the  
CC solubility of the first domain on its own, or (c) both to translocate  
CC the protein into a cell and to increase the solubility of the protein  
CC compared to the solubility of the first domain on its own, the protein  
CC being free of clostridial neurotoxin (CN) and free of CN precursor that  
CC can be converted into toxin by proteolytic action. The recombinant  
CC proteins can be used as therapeutic agents for targeting cells  
CC expressing a relevant substrate. The products can also be used as  
CC immunogens and as non-toxic standards for the assessment and development  
CC of in vitro assays for the detection of functional botulinum or tetanus  
CC neurotoxins either in foodstuffs or in environmental samples.

XX Sequence 3509 BP; 1468 A; 340 C; 548 G; 1153 T; 0 other;

Query Match 39.0%; Score 522.4; DB 19; Length 3509;  
Best Local Similarity 71.7%; Pred. No. 4.9e-134;  
Matches 685; Conservative 0; Mismatches 271; Indels 0; Gaps 0;

QY 10 atggccaaacaaatacaattccgaatccctgaacataatctgaacctgcggtacaaa 69  
Db 2554 atgtttaataataatagcgaatttttaataataattattcttaatttaagataaag 2613

QY 70 gacaaactctgactgctgtctgttacggtgtctaaagttgaagtatacacagtggt 129  
Db 2614 gataataatttaagatttatcaggatatggggcaaggtagaggtatgatggagtc 2673

QY 130 gaactgaatgacaagaaccagttcaaacctgacctctccgctaaactctaaagtcggtgt 189  
Db 2674 gagcttaagtataaaatacaatttaataactagttcagcaaaatagtaagtagtg 2733

QY 190 actcagaatcagaacatcatcttcaactccgttattctctggactctctgttctctctgg 249  
Db 2734 actcaaatcagaatcatcatatttaagtgtgttctctgttttttagcgttttgg 2793

QY 250 attcgtatcccgaaatacaagaacgacggttatccagaattatcaccaatgaatacacc 309  
Db 2794 ataagaatacctaaataaagaatgatggtatcacaaatattatcataatgaataca 2853

QY 310 atcatcaactgcgtgaagaataaactctgtgttggaagatctccatccggttaacggtatc 369  
Db 2854 ataataatgtatgaaaaataattcggcgtggaataatctattaggggtaaataggata 2913

QY 370 atctggactctgatcatatcatcaacggtgaagaccaaatctgtattcttctogaatacaacatc 429  
Db 2914 atatggactttaattgatataaatggaaaaaccacaaatcggtatttttttgataatacata 2973

QY 430 cgtgaagacatctctgaatacatcaatcgctgtgttctctcgttaccatcaccaacatcg 489  
Db 2974 agagaagatatcatcagagtataataatagatggtttttttgtaactattactaataattg 3033

QY 490 aacaatgctaaatctacatcaacggttaaacctggaatctaataaccgacatcaaaagacatc 549  
DB 3034 aataacgctaaaatttataatgaatgtaagctagaatcaaatcacagattataaagatata 3093  
QY 550 cgtgaagttatcgttaacggtgaataatcatcttcaaacctggagcgttgacatcgatcgacc 609  
DB 3094 agagaagttattgctaatggtgaataatatttaattagattggtgatagatagaaca 3153  
QY 610 cagttcatctgataaaatcacttccatcttcaacacccgaaactgctcagtcacaaatc 669  
DB 3154 caattatttgatgaataattattcagttatttataacaggaatttaagtcacaaatatt 3213  
QY 670 gaagaacggtacaaatccatcttactccgaataacactgaaagacttctgggttaactcg 729  
DB 3214 gaagaagatataaaattcaatcatagcgaattttaaaagatttttggggaaatcct 3273  
QY 730 ctgattgtacaaagaataactatatttcaatgttgtaacgtgtgaacaaagactttacatcaa 789  
DB 3274 ttaattgtacaataaagaattatttatgtttaatgctggggaataaaattcatattataa 3333  
QY 790 ctgaagaagaactcccggttggtgaatccctgactcgttccaaataacacagaactct 849  
DB 3334 ctgaagaagaactcccggttggtgaatccctgactcgttccaaataacacagaactct 3393  
QY 850 aaatcatcaactaacgcgacactgtacatcgttgaaagtctcatcgcgtgcgaactct 909  
DB 3394 aaatataaaattatagatttatatttggaagaaatttattataagaagaagtca 3453  
QY 910 aactctcagtcacaaatgatgacatcgtaacgttaacgaagaactacatctacctgga 965  
DB 3454 aattctcatctataaagatgatagattagttagaagaagaagattatattatctaga 3509

## RESULT 8

AAA54491  
ID AAA54491 standard; DNA; 1368 BP.

XX

AC AAA54491;

XX 11-APR-2001 (first entry)

XX Botulinum toxin heavy chain C-terminal coding sequence (serotype G).

KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;

KW recombinant vector; antigen; immune response; vaccine; bacterium;

XX infection; ds.

XX Synthetic.

OS Clostridium botulinum.

XX Key Location/Qualifiers

FH 10..1359

FT CDS

FT /\*tag= a

FT /product= H<sub>2</sub>C peptide fragment

XX WO200067700-A2.

XX 16-NOV-2000.

XX 12-MAY-2000; 2000WO-US12890.

XX 12-MAY-1999; 99US-0133865.

PR 12-MAY-1999; 99US-0133866.

PR 12-MAY-1999; 99US-0133867.

PR 12-MAY-1999; 99US-0133868.

PR 12-MAY-1999; 99US-0133869.

PR 12-MAY-1999; 99US-0133873.

PR 29-JUL-1999; 99US-0146192.

XX (USSA ) US ARMY MEDICAL RES &amp; MATERIAL COMMAND.

XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;

XX

DR WPI; 2001-016048/02.  
DR P-PSDB; AAB04167.

XX New nucleic acids encoding the carboxy- or amino-terminal portions of  
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
PT vaccine against botulism

XX Claim 2; Fig 10a; 73pp; English.

XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
CC chain and then posttranslationally nicked, forming a dichain  
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
CC remain linked by a disulfide bond. Nucleic acids encoding the  
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
CC expression vectors and expressed in transformed cells to produce  
CC peptide antigens useful for eliciting an immune response to give  
CC protective immunity against botulinum neurotoxin, which causes  
CC botulism. The nucleic acids are expressible in a recombinant  
CC organisms such as Escherichia coli or Pichia pastoris. The use  
CC of recombinant nucleic acids are advantageous since it eliminates  
CC the need to culture large quantities of hazardous toxin-producing  
CC bacterium. Production yield from the genetically engineered product  
CC is also high and cost of production is lower. The nucleic acids can  
CC be derived from Clostridium botulinum serotypes A-G.

XX Sequence 1368 BP; 366 A; 445 C; 267 G; 290 T; 0 Other;

Query Match 31.8%; Score 426.8; DB 22; Length 1368;  
Best Local Similarity 60.5%; Pred. No. 8.7e-108;  
Matches 818; Conservative 0; Mismatches 507; Indels 27; Gaps 6;

QY 5 tcacgatggccaaataacaaattccgaaatcctgaaataatcctgaacaaatcctcctgaacctcggtt 64  
DB 29 tccaggtcttcaacactacatctccacatctctcccaacgcacatcctgctcctgctcct 88  
QY 65 acaaaagacaaacaaatcgtatcgtctgtgttacggtgtctaaagttaagtatacagacg 124  
DB 89 accgtggt 148  
QY 125 gtgtggaactgaatgac-----aagaacagttcaaaactgacctcttctcgttaactcta 178  
DB 149 acgtctcttcaacgacatcggttaacggtcagttcaagctgaacaaactcgcgaactcca 208  
QY 179 agatccgtgttactcagaatacagaacatcatcttcaactccgtattctcgtggaactctcg 238  
DB 209 acatcacgcgccaccagtcacaaagtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 268  
QY 239 ttctcttctggttctgttccgaaataacaaagacggttatccagaattacatccaca 298  
DB 269 tcaacttctggttcgttaccacaaagatacaacaaacacacacacacacacacacacacac 328  
QY 299 atgaatacaccatcatcaactcgaatgaagaataactctgtgtgtgtgtgtgtgtgtgtgtgtgt 358  
DB 329 acgagtacaccatcatctctgtatcaagaacgactcgcgtgtgtgtgtgtgtgtgtgtgtgtgt 388  
QY 359 gtaaccgtatcatctggaactgtgatcatcaacggttaagacacaaatctgtattcttctg 418  
DB 389 gaaacggtatcatctggaacctgtatcgatcgatcaacgccaagtcacagtcacatcttctg 448  
QY 419 aatacaacatccgtgaagacatctctgtgaatacatcaaatcgtgtgtgtgtgtgtgtgtgtgtgt 478  
DB 449 agtactccatcaaggacacacatctcgcgactacatacaacaaagtgtgtgtgtgtgtgtgtgtgt 508  
QY 479 ccaataaac---ctgaacaatgtctaaatctacatcaacggttaaacctggaactcttaatacgg 535  
DB 509 ccaacgacgctgt 568  
QY 536 acatcaaaagacatccgtgaagtattatcgcttaacggtgaataatcatcttcaaacctggaacggt 595  
DB 569 agatctggaacctggaccgtatcaactctccaaacacacacacacacacacacacacacacacacac 628







Db	260	tcaactctctgggttcogttaccoccaaagttacaacaacagcatccagactcacttcacgtcaga	319
Qy	299	atgaatacaccatcatcaactgcgatgaagaaataactctggttggaagatctccatccgcg	358
Db	320	acgagttacacatcatctctctgtatcaagaaacgactcgggttggaaggtctccatcaagg	379
Qy	359	gtaacggtatcatcttggaactctgatogtatatacaacggttaagacaaaatctgttatcttcg	418
Db	380	gaacggtatcatctggaccctgatcgacgtcaacgccaagtccaagtcccatctctcttcg	439
Qy	419	aatacaacatccgtgaagacatctctgtaatacatcaatcgtggtctctctgttaccatca	478
Db	440	agttactccaacaggacaacatctctcgactactacaacaaagtggcttccatccaccatca	499
Qy	479	ccaataac---ctgaacaatgtcaaaatcatcatcaacacggttaaacctggaatcttaataccg	535
Db	500	ccaacgacctctgggttaacgccaacatctcatcaacggttccctgaagaagtcgcgaga	559
Qy	536	acatcaaaagacatccgtgaagttatctgttaaacggtgaaatcatcttcaaacctggaacggtg	595
Db	560	agatcttgaaacctggaccgtatccaactctcaacgacatcgacttcaagctgatacact	619
Qy	596	acatcgatcgaccaggttcactctgatgaaatacttctccatcttcaaacgcgaactgt	655
Db	620	gtaccgaacccaaggttcgtctggtgaaggaacttcaacatctctcgtctgtagtctga	679
Qy	656	ctcagtcacaatctcgaagaacggtcaagaatccagctcttactctcgaa tacttgaaagact	715
Db	680	acgccaccgaggtctcctccctgactggatccagctctctccacaacacctggaagact	739
Qy	716	tctggggtaatccgctgtatgtacaacaagaatactataatgttccaatgctgggttaacaaga	775
Db	740	tctggggaacccactcgtttacgacacocagtagtactcctgttcaaccagggtagcaga	799
Qy	776	actcttatcaaacactgaagaaagactctcgggttggtggaactcctgactgttccaaat	835
Db	800	acatctaca tcaagttacttctccaaggctccatcgggtgagacgcgcctctgtaaccaact	859
Qy	836	acacaggaactctaaatacatcaactacgcgacctgtacatcgggtgaaagttcatca	895
Db	860	tcaacaacgcgcg----catcaactaccagaacctgtacctgggtctgcgttttcatca	913
Qy	896	tcogtcgcaaaatctaactc---tcagttcaatcaatgtatgacatcgtacgttaagaaagact	952
Db	914	tcagaagggcctccaactccgttaacatcaacaacgacaacatcgtccgtgaggggtgact	973
Qy	953	acatctacctggactcttcaacct---gaatcaggaatggcgtgtatacacctacaagt	1009
Db	974	acatctaccttgaacatcgcagacaatctcgcagagtcctacgtgtctatcgtctcgtgtca	1033
Qy	1010	acttcaagaagaagaaagaaagcttttctcgtcccgatctctgtatctccgaacgaactct	1069
Db	1034	actccaaggagatccgaaccacgcttctcgtggcccaatcaacgagaccctactctct	1093
Qy	1070	acaaacacatccagatccaagaatacaagactactctgcataccgacactactcttgacagctgtgt	1129
Db	1094	acgacgtcctcgtcagatcagaagaagtactacgagaagaccacatacaactgtcagatctcgt	1153
Qy	1130	tcaagaagaatgaagaatctactgcagaaatcgggtctgatcgggtatccacogtttctacg	1189
Db	1154	gcgagaggacaccaagacactctcgagctgttcgttgcgtatcggttaagtctg-----tcaagg	1207
Qy	1190	aactcgtgtactgtattcgaagaatacaagactactctgcatactccaatggtactcga	1249
Db	1208	actacggttacgctcgtggacactcagacaactactctgtatctccacgtggcacctcgc	1267
Qy	1250	aggaagttaaacgcaaacccgtacaacctgaactgggttgccaattggcagttcatcccca	1309
Db	1268	gtcgtatctccgagaacatcaacaagctgcgtctgggagtgaactggcagttcatcccg	1327
Qy	1310	aagacgaaggttgaccga 1328	
Db	1328	tcacacgaaggttgaccga 1346	

RESULT 10	
AAA54483	
ID AAA54483 standard; DNA; 1323 BP.	
XX AC	
XX AC AAA54483;	
XX DT	
DT 11-APR-2001 (first entry)	
XX BA	Botulin toxin heavy chain C-terminal coding sequence
DE BA	
XX BA	Botulism; toxin; neurotoxin; heavy chain; recombinant vector; antigen; immune response; infection; ds.
KW KW	
KW KW	
OS OS	Synthetic.
OS OS	Clostridium botulinum.
XX XX	
FH FH	Key Location/Qualifiers
FT FT	13..1317
FT FT	/tag= a
FT FT	/product= H_C peptide fragment
XX XX	
PN PN	WO200067700-A2.
XX XX	
PD PD	16-NOV-2000.
XX XX	
PF PF	12-MAY-2000; 2000WO-US12890.
XX XX	
PR PR	12-MAY-1999; 99US-0133865.
PR PR	12-MAY-1999; 99US-0133866.
PR PR	12-MAY-1999; 99US-0133867.
PR PR	12-MAY-1999; 99US-0133868.
PR PR	12-MAY-1999; 99US-0133869.
PR PR	12-MAY-1999; 99US-0133873.
PR PR	29-JUL-1999; 99US-0146192.
XX XX	(USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND
PA PA	
XX XX	Smith LA, Byrne WP, Middlebrook JL, Lapenotier
PI PI	
XX XX	
DR DR	WPI; 2001-016048/02.
DR DR	P-PSDB; AAB04089.
XX XX	
PT PT	New nucleic acids encoding the carboxy- or amino-
PT PT	the heavy chain of botulinum neurotoxin of serotype
PT PT	vaccine against botulism
XX XX	
PS PS	Disclosure; Fig 2a; 73pp; English.
XX XX	
CC CC	Botulin neurotoxins are translated as a single polypeptide
CC CC	chain and then posttranslationally nicked, forming two chains
CC CC	consisting of a 100 kDa heavy chain and a 50 kDa light chain
CC CC	remain linked by a disulfide bond. Nucleic acid sequences encoding
CC CC	carboxy-terminal (HC) or amino-terminal (HN) portions of the
CC CC	chain of botulinum neurotoxin (BoNT) can be used to generate
CC CC	expression vectors and expressed in transformed cells to produce
CC CC	peptide antigens useful for eliciting an immune response
CC CC	protective immunity against botulinum neurotoxin.
CC CC	botulism. The nucleic acids are expressible in prokaryotic
CC CC	organisms such as Escherichia coli or Pichia pastoris.
CC CC	of recombinant nucleic acids are advantageous for the production
CC CC	the need to culture large quantities of hazardous organisms
CC CC	bacterium. Production yield from the genetically modified
CC CC	is also high and cost of production is lower than that of
CC CC	be derived from Clostridium botulinum serotypes.
XX XX	
SQ SQ	Sequence 1323 BP; 404 A; 334 C; 242 G; 343 T; 343 C; 343 G; 343

Query Match 19.6%; Score 263; DB 22; Length 1323;  
Best Local Similarity 53.9%; Pred. No. 1.7e-62;

Matches 639; Conservative 0; Mismatches 525; Indels 21; Gaps 4;

```
Qy 15 caacaaatacattcgaatacctgaacataatatactgaacacgtggttacaagaaca 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24 cactgaatacatcaagaaacatcatcaatactcactgaacacgtggttacaagaaca 83
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 75 caactgatgatctgctggttacggtgctaaagtgaagtatacagcaggtgtt---- 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 tcacctgatgcagcgtctctgcgtacgtgtccaaatacaacatcggttctaagttaactt 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 130 -gaactgaatgacaagaacacgttcaactgaactcttccgctaaacttaagatccgtgt 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 cgatccgatgacaagaatcagatccagctgttcaactgtgaactctccaaaatcgaagt 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 189 tactcagaatacagaacatactctcaactcgttattctcgtgaacttctcttctctctg 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 204 tatectgaagaatgctatctatatacactctatgtacgaaaacttctccacctctctg 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 249 gattcgtatccgaaatacagaacacggtgtatccagaattatccacatagaatacac 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 264 gatccgtatccgaaatacacttaactcc-----atctctgaaacatgaatacac 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 309 catcataactgcatgaagaataactctgttggaagatctccatccgcggtgaacgtat 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 catcataactgcatgaagaataactctgttggaagatctctgaactacggtgaat 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 369 catctgactctgatcatcaacggtgaagacaaatctgtattcttctgaatacaaat 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 375 catctgactctgaggaactcaggaactcgaacagcgtgtgttctcaataactctca 434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 429 ccgtgaagacatctctgaatacatcaactcgtgttctgttaccatcccaataa--- 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 gatgatacaactctgatacatcaactcgtgtgtgttaccatcccaataaactg 494
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 486 cctgaacaaatgctaaatactatcaacacggttaactggaatctaataccgacatcaaga 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 495 tctgaataactcgaatactatcaacacggtgtgtgtgtgtgtgtgtgtgtgtgtgt 554
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 546 calccgtgaagtattcgtcaacacggtgaatactctcaactcgaacggtgacatcgatg 605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 555 tctggtgcaacatccacgtcttcaataacatcatgttcaactggtgacgtgtgacac 614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 606 taccaggttcatctggtgaaatactctccatcttcaacacacgaaactgtctcagtcaca 665
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 615 tcacccgtatcatctggtatcaataacttcaactctgttcgacaaagaaactgaacgaaaga 674
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 666 tatagaagaacggttacaagatccagcttactccgaataactgaagaacactctggtgaa 725
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 aatcaagaacgtgtacgacacacgttccaaatctgtgtatctgaaagacttctgggtga 734
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 726 tccgctgatgtacaacaaagaatactatgttcaatgctgtggttaacaaagactcttactat 785
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 735 ctacctgagtacgacaaacggtactacaatgctgaatctgtacgacatccgaaacaaactgt 794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 786 caaactgaagaagaactctccggtgtgtgaaatctgactcgttccaaatacaacacgaa 845
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 795 tgacgtcaacaatgtaggtatccggttaccatgacctgaaggtccgcggtgtctgt 854
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 846 ctctaatacatcaactacacgcgacactgtacatcgttgaaagtctcatccgtcgcaa 905
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 855 tatgactaccaacatctacgtgaactctccgtgtacccgtgtggttaccacaaatcatcatca 914
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 906 atctaactctcagtcctcatcatgatgatcgtacgtacgtgaagaagactacatctaccctgga 965
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 915 gaaatagcgtctgtgtaacaaggaacaatacgttccgcaacaatgatcgtgtatacatcaa 974
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 966 ctcttcaactgaatcaggaaatggcgtgtatatacacatcaagaacttctcaagaagaaga 1025
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 975 tgttgtagtaagaacaaagaataacccgtctggtaccacaaatgcttctcaggtgtgtaga 1034
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1026 agaaaagcttctcgtcgtccgtatctgtattccgaagaaactctacaacacacacacagat 1085
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1035 aaagatctgtctgctgtgaaatcccgacgttggtaactgtctcaggtaggtgtaat 1094
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Qy 1086 caaagaatagcagcaacagccgacctactctgtccagctgtgttcaagaagaatgaaga 1145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1095 gaaatcccaagacgacccaaggta---tcactaaacaatgcaaaatgaatctgcaggacaa 1151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1146 atctactgacgaaatcggtctgctgacgttatccacccttcttacga 1190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1152 caatggttaacgatacgtttctcatcgtttccaccaggttcaacaa 1196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 11  
AAA54484

ID AAA54484 standard; DNA; 1326 BP.

XX AC AAA54484;

XX DT 11-APR-2001 (first entry)

XX DE Botulism toxin heavy chain C-terminal coding sequence (serotype A).

XX KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;

XX KW recombinant vector; antigen; immune response; vaccine; bacterium;

XX KW infection; ds.

XX OS Synthetic.

XX OS Clostridium botulinum.

XX Key Location/Qualifiers

FT CDS 13..1320

FT /\*tag= a

FT /product= H\_C peptide fragment

XX PN WO200067700-A2.

XX PD 16-NOV-2000.

XX PF 12-MAY-2000; 2000WO-US12890.

XX PR 12-MAY-1999; 99US-0133865.

XX PR 12-MAY-1999; 99US-0133866.

XX PR 12-MAY-1999; 99US-0133867.

XX PR 12-MAY-1999; 99US-0133868.

XX PR 12-MAY-1999; 99US-0133868.

XX PR 12-MAY-1999; 99US-0133873.

XX PR 29-JUL-1999; 99US-0146192.

XX PA (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.

XX PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;

XX DR WPI: 2001-016048/02.

XX DR P-PSDB; AAB04090.

XX PT New nucleic acids encoding the carboxy- or amino-terminal portions of  
the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
vaccine against botulism

XX PS Disclosure; Fig 3a; 73pp; English.

XX CC Botulism neurotoxins are translated as a single 150 kDa polypeptide  
chain and then posttranslationally nicked, forming a di-chain  
consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
remain linked by a disulfide bond. Nucleic acids encoding the  
carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
chain of botulinum neurotoxin (BoNT) can be used in recombinant  
expression vectors and expressed in transformed cells to produce  
peptide antigens useful for eliciting an immune response to give  
protective immunity against botulinum neurotoxin, which causes  
botulism. The nucleic acids are expressible in a recombinant  
organism such as *Escherichia coli* or *Pichia pastoris*. The use  
of recombinant nucleic acids are advantageous since it eliminates  
the need to culture large quantities of hazardous toxin-producing  
bacterium. Production yield from the genetically engineered product

CC is also high and cost of production is lower. The nucleic acids can be derived from *Clostridium botulinum* serotypes A-G.

XX Sequence 1326 BP; 404 A; 336 C; 243 G; 343 T; 0 other;

Query Match	19.6%	Score 263	DB 17	Length 1330
Best Local Similarity	53.9%	Pred. No. 1.7e-62		
Matches 639	Conservative 0	Mismatches 525	Indels 21	Gaps 4
QY	15	caacaaatacaatcccgaaatccgaaacaatatcatctgaacctgcgttaacaagacaa	74	
DB	24	cactgaatacatcaagaacacacaaatcatctcatctgaacctgcgttaacaagacaa	83	
QY	75	caatctgacgatctgtctggttcggtgcgttaaggttgaagtatacgcggtgtt----	129	
DB	84	taactgatgcgctgtctgcgtacgtctccaaatcaaatcggttctaaagttaactt	143	
QY	130	-gaactgaatgacaagaacaaagttcaaaactgacgtcttcctgaacttaagatccgtgt	188	
DB	144	cgatccgatgcagaagaatcagatcccggtgttccaatcgtgaattctccaaatcgaagt	203	
QY	189	tactcagaatcagaaacatcatctccaaactcgtattcctggaacttctgtttcctctgt	248	
DB	204	tatcctgaagaatgctatcgtatatacaactctatgtacgaaaacttctccacctctctgt	263	
QY	249	gattcgtatcccgaaatataagaacgcggttatccgaattatcatccaatgaataaac	308	
DB	264	gatccgtatcccgaaatattctcaactcc-----atctctctgaacaatgaatacac	314	
QY	309	catcataactgcgatgaagaataaactcgtgttggaagtctccatccgcggtgaacctat	368	
DB	315	catcatcaactgcgatgaagaataaactcgtgttggaagtatctctgaactacggtgaat	374	
QY	369	catctggaactctgatcatatacaacggttaagaccaaactctgtattcttcgaatacaaat	428	
DB	375	catctggaactctgatcatatacaacggttaagaccaaactctgtattcttcgaatacaaat	434	
QY	429	ccgtgaagacatctgaatacatcaatcgaactcgtgttcttcgaactcaaccaataa--	485	
DB	435	gatgatacaactctgatcatatacaacggttaagaccaaactctgtattcttcgaatacaaat	494	
QY	486	cctgaacaatgctaaaaatctacatacaacggttaaaactggaatcctaataccgacataaaga	545	
DB	495	tctgaataactccaaatactacatacaacggttaaaactggaatcctaataccgacataaaga	554	
QY	546	catccgtgaagtatcgttaaacggtgaatacatcttcaaaactggaacggtgaactgcgtgcg	605	
DB	555	tctggtgcaacatccacgtcttaataaacaatcgttcaaaactggaacggttgcgtgcac	614	
QY	606	taccaggttcaatctggatgaataactctccatcttcaacacggaactgtctcgtaccaa	665	
DB	615	tcaccgttcaatctggatgaataactctccatcttcaacacggaactgtctcgtaccaa	674	
QY	666	tatogaagaacggttacaagatccaagtcttactccgaataactgaagaacttctgggttaa	725	
DB	675	aatacagaactgtacgacacacggtccaactctcgttatcctgaagaacttctgggttga	734	
QY	726	tcogctgatgtacacaagaataactatattgttcaatgctggttaacaagaactcttaact	785	
DB	735	ctacctgcgtacgacaaaacggtactacatgctggaatctgtacgataccgaataactcgt	794	
QY	786	caaaactgaagaagactctccggttggtgaaatcctgactcgttccaaatacaaacacgaa	845	
DB	795	tgaactcaacaatgaggtatccggtgttaactgatactgaaaggtcccggtgttctgt	854	
QY	846	ctctaaatacatcaactaccgcgacgtgatacgtgtgaaatcctgactcgttccaaatacaaacacgaa	905	
DB	855	tatgactaccaacaactcacctgaactctccctgataccggtgttaccaaaatcatcatcaa	914	
QY	906	atctaatctcagttccatcaatgatgacatcgttacgtacgtaaaagaagactacatctacctgga	965	
DB	915	gaaatacgcgtctggttaacaaggaacatacgttgcgaacaatgatcgtgtatacatcaa	974	
QY	966	cttcttcaacctgaatcagggaatggtgcgttatacaacctcaacgtattcttcaagaagaaga	1025	
DB	975	tgtgtgtagttaaagaacaaagatacgtcgttgcgttaccacatgcttctcagcgtgtgtgaga	1034	





CC this vector results in the production of large amounts of a protein  
CC encoded by a sequence cloned into the replicon. The constructs are used  
CC to produce vaccines against botulism. The proteins can also be used as  
CC diagnostic tools for the diagnosis of botulism. The transformed host  
CC cells can be used to analyse the effectiveness of drugs and agents which  
CC inhibit toxin effects. The vaccine currently used against botulism is  
CC dangerous and expensive to produce, and contains formalin, which is very  
CC painful for the recipient. Also, the vaccine is incomplete, in that only  
CC 5 of the 7 serotypes are represented in the formulation. The novel  
CC vaccine overcomes these problems, as it is easily purified, and  
CC available in large quantities. It is also expressed in the lymph nodes  
CC for a better immune response. Sequences A4287212-287217 represent  
CC synthetic DNA sequences encoding BoNT Hc fragments used in the present  
CC invention. These were optimised for codon usage for expression in yeast.  
XX

Sequence 1338 BP; 401 A; 342 C; 249 G; 346 T; 0 other;

Query Match 19.6%; Score 263; DB 21; Length 1338;  
Best Local Similarity 53.9%; Pred. No. 1.8e-62;  
Matches 639; Conservative 0; Mismatches 525; Indels 21; Gaps 4;

Qy	15	caacaaatacaatccgaatactcgaacataatcatctcgaacacggtggttacaagaacaa	74
Db	32	cactgaatacatcaagaacatcatcaatcacctccatctgaacctgcgctacgaatccaa	91
Qy	75	caatctgatcgtctgtgttaccggtgttaagtgtgaagtatacagcgggttt-----	129
Db	92	tcacctgatcgtctgtctcgtcgtcgtctccaaatacaacatcgtgttetaaagttaactt	151
Qy	130	-gaactgaatgacaaacaccgtttcaactgacctcttccgtcaacttaagatccggtat	188
Db	152	cgatccgatcgaagaatcagatccagctgttcaatctggaatcttccaaatcgaagt	211
Qy	189	tactcagaatacagaacatcatcttcaactcogtattctcgtgaactctctgtttctcttg	248
Db	212	tatctgaagaatgctatcgtatatacaactctatgtacgaacacttctccactcctctgt	271
Qy	249	gattcgtatccgaataacagaacgacggtatccagaattacatccacaaatgaatacac	308
Db	272	gatccgtatccgaataacttcaactcc-----atctctctgaacaaatgaatacac	322
Qy	309	catcatcaactcgtgaagaataactctgttggaaagtatccatcccggtgaaccggtat	368
Db	323	catcatcaactcgtgaagaataactctgttggaaagtatctcgaactcaggtgaaat	382
Qy	369	catctggactctgatgatatacagcgtgaagacaaatctgtattcttcgaatacaacat	428
Db	383	catctggactctgagdacactcaggaataatcaaacagcgtgtgtattcaataactctca	442
Qy	429	ccgtgaagacatctctgaatacatcaactcgtgttggaaagtatcaataccgacatcaaga	485
Db	443	gatgatacaactctctgactacatacaactcgtgtggtatctcttaccatccacaataa	502
Qy	486	cctgaacaactgtaaaatctcatcaacggttaactggaatctaaatccgacatcaaga	545
Db	503	tctgaataactccaaatctcatcaacggtcgtgtgatctcttaccatccacaatacgc	562
Qy	546	catccgtgaagtattcgttaacggtgaataatcttcaaacgtgacggtgacatcgatcg	605
Db	563	tctgggttaacatccacgctcttcaatacatcatgttcaaacgtgacggtgtcgtgacac	622
Qy	606	taccagttcatctggatgaataactctctccatcttccaaacggaactgtctcagttcaa	665
Db	623	tcacgctacatctctggatccaaatctcaatctgttcgacaaagaactgaacgaaaaaga	682
Qy	666	tatcgaagaacggttacagatccagtttactccgaatacactgaaagacttctgggttaa	725
Db	683	aatcaaaagacctgtacgacaaacggtccaattctgtgtatctcgtgaagacttctgggtga	742
Qy	726	tcgcgtgatgtacaagaagaataactatattgttcaatgctgggttaacaagaactcttaac	785
Db	743	ctaactgcgtacgacaaacggtactactatctgtgaatctgtacgatccgacaataacgt	802

Search completed: September 2, 2002, 17:08:55  
Job time: 4037 sec

Qy	786	caaaactgaagaaagactctccggttggtggaatcctgaactggttccaaatacaaccagaa	845
Db	803	tgagctcaacaatgtaggtatcccggttcatctgacctgaaggtccgcgtggttctgt	862
Qy	846	ctctaaaatacatcaactaccgcgacctgtatcatcgtgtaaaagtttcatcaccgctcgcaa	905
Db	863	tatgataccaacatctcactgaactcttccctgtacccgtggtaccgaattcatcatcaa	922
Qy	906	atctaaactctcagtcocatcaatgatgatcgtacgtacgttaaaagaactacatctacctga	965
Db	923	gaaatacgcgtctgtgtaacaaggacaaatcgtttcgcgaacaatgatcgtgtatatacatcaa	982
Qy	966	cttcttcaacctgaatcagggaatggtgtgtatatacacctacacgtacttcaagaagaaga	1025
Db	983	tggtgtagttaaagaacaaagataccgctcgtggtaccacatgcttctcaccaggtgtgtaga	1042
Qy	1026	agaaaagcttttctcgtcgcgtatctctgattccgcgacgaacatctacaacaccatccagat	1085
Db	1043	aaagatctgtctgtctcgtggaatcccggaacgttggtaatctgtctcaggttagttgtaat	1102
Qy	1086	caaagaatacagacgaacagcgaactactcttgcagctgtgttcaagaagaatgaaga	1145
Db	1103	gaaatccaagaacgaccagggta---tcaactaacaatgcaaaatgaatctgcaggacaa	1159
Qy	1146	atctactgacgaataatcgtgtcgtgatccggtatccacggttcttctacga	1190
Db	1160	caatggttaacgatcgtgttccatcgtgttccaccaggttccaccaggttcaacaa	1204





OM of: US-09-910-186a-8 to: GenEmbl:\* out\_format : pfs

Date: Sep 2, 2002 3:59 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=framet-p2n.model -DEV=xlh  
-O=/cn2.1/USPTO.spool/US09910186/runat_29082002_134725_15992/app_query.fasta_1.505  
-DB=GenEmbl -QFWT=fastap -SUFFIX=rge -CAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFWT=pfs -NORM=ext -HEARSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09910186_@CGN1_1_4342  
-NCPUP=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30  
-NO_XLUPY -WAIT -THREADS=1
```

Search information block:

Query: US-09-910-186a-8

Query length: 440

Database: GenEmbl.\*

Database sequences: 1797656

Database length: 187333701

Search time (sec): 1881.260000

score\_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb_ba:CLOBOTB	+	2340.00	2577.97	3.1e-135	4041	M81186 Clostridium botulinum n
gb_ba:AF295926	+	2231.00	2458.15	1.5e-128	3876	AF295926 Clostridium botulinum
gb_ba:AF300465	+	2229.00	2455.94	2.0e-128	3876	AF300465 Clostridium botulinum
gb_ba:AF300466	+	2229.00	2455.94	2.0e-128	3876	AF300466 Clostridium botulinum
gb_ba:CBY13630	+	2223.00	2441.57	1.2e-127	11170	Y13630 Clostridium botulinum
gb_ba:AF300467	+	2210.00	2435.01	2.9e-127	3869	AF300467 Clostridium botulinum
gb_ba:AF300468	+	2210.00	2435.01	2.9e-127	3869	AF300468 Clostridium botulinum
gb_ba:AF300469	+	2210.00	2435.01	2.9e-127	3869	AF300469 Clostridium botulinum
gb_ba:CBBOTB	+	2099.00	2312.35	2.0e-120	4051	AF1343 C.botulinum type B gene
gb_ba:CB024628	+	2097.00	2318.32	9.1e-121	1326	AJ242628 Clostridium botulinum
gb_ba:PA69701	+	1667.00	1837.30	5.6e-94	3509	AF59701 Sequence 19 from Patent
gb_ba:BD009886	+	1667.00	1837.30	5.6e-94	3509	BD009886 Recombinant toxin fra
gb_ba:CBNTG	+	1168.00	1287.07	2.5e-63	3937	X74162 C.botulinum Bont/G gene
gb_ba:CBNTNHTB	+	1103.00	1210.81	4.4e-59	6862	X78749 C.botulinum NTH gene &
gb_pat:AR000029	+	831.50	923.61	4.4e-43	1330	AR000029 Sequence 22 from pate
gb_pat:AR169140	+	831.50	923.61	4.4e-43	1330	AR169140 Sequence 22 from pate
gb_pat:AX036243	+	831.50	923.61	4.4e-43	1330	AX036243 Sequence 22 from Pate
gb_sy:XXU22962	+	831.50	923.57	4.4e-43	1338	U22962 Synthetic botulinum neu
gb_pat:AR000030	+	831.50	923.23	4.6e-43	1402	AR000030 Sequence 25 from pate
gb_pat:AR169141	+	831.50	923.23	4.6e-43	1402	AR169141 Sequence 25 from pate
gb_pat:AX036246	+	831.50	923.23	4.6e-43	1402	AX036246 Sequence 25 from Pate
gb_pat:AR000031	+	831.50	915.75	1.2e-42	3891	AR000031 Sequence 27 from pate
gb_pat:AR169142	+	831.50	915.75	1.2e-42	3891	AR169142 Sequence 27 from pate
gb_pat:AX036248	+	831.50	915.75	1.2e-42	3891	AX036248 Sequence 27 from Pate
gb_ba:CBBOTAG	+	831.50	915.03	1.5e-42	4292	X52066 Clostridium botulinum B
gb_ba:CLOBOTB	+	831.50	914.16	1.5e-42	4835	M30196 C.botulinum neurotoxin
gb_ba:CBBOTF	+	778.00	856.45	2.4e-39	4073	X68262 C.barati gene for type
gb_ba:CBNTQXA	+	777.50	855.91	2.6e-39	4067	X73423 C.botulinum gene for in
gb_ba:CBBPOT	+	746.00	830.88	6.4e-38	1084	X70817 C.botulinum (strain pro
gb_pat:AS8945	+	732.00	814.16	5.5e-37	1293	AS8945 Sequence 5 from Patent
gb_sy:AF251281	+	732.00	814.13	5.5e-37	1299	AF251281 Synthetic construct
gb_pat:AS8946	+	732.00	814.05	5.6e-37	1313	AS8946 Sequence 6 from Patent
gb_ba:CLOBOT	+	732.00	805.54	1.7e-36	4199	L35496 Clostridium botulinum B
gb_ba:CBBOTFG	+	732.00	805.52	1.7e-36	4209	X81714 C.botulinum bont/f gene
gb_ba:CBY13631	+	723.00	790.32	1.2e-35	9325	Y13631 Clostridium botulinum B
gb_ba:CLOBOTROX	+	723.00	795.56	6.0e-36	4234	M92906 Clostridium botulinum B
gb_ba:CBB2129	+	718.00	800.03	3.4e-36	1084	X70819 C.botulinum (strain non
gb_ba:CBBEKLUND	+	718.00	800.03	3.4e-36	1084	X70819 C.botulinum (strain non
gb_ba:CBNTTE	+	692.50	762.30	4.3e-34	4030	X62683 C.botulinum bontE gene f
gb_ba:CBNEUTOXE	+	692.00	761.78	4.6e-34	4017	X62089 C.botulinum gene for bd

gb\_pat:AX088262 + 692.00 761.78 4.6e-34 4017 ! AX088262 Sequence 10 from P  
gb\_ba:AB037704 + 683.00 752.20 1.6e-33 3835 ! AB037704 Clostridium butyri  
gb\_ba:AB037705 + 683.00 752.20 1.6e-33 3835 ! AB037705 Clostridium butyri  
gb\_ba:AB037706 + 683.00 752.20 1.6e-33 3835 ! AB037706 Clostridium butyri  
gb\_ba:AB037707 + 683.00 752.20 1.6e-33 3835 ! AB037707 Clostridium butyri

seq\_name: gb\_ba:CLOBOTB

seq documentation block:

LOCUS CLOBOTB 4041 bp DNA linear BCT 26-APR-1993  
DEFINITION Clostridium botulinum neurotoxin type B (bont) gene, complete cds.  
ACCESSION M81186  
VERSION M81186.1 GI:144734  
KEYWORDS bont gene; neurotoxin type B.  
SOURCE Clostridium botulinum DNA.  
ORGANISM Clostridium botulinum  
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
Clostridium.

REFERENCE 1 (bases 1 to 4041)  
AUTHORS Whelan,S.M., Elmore,M.J., Bodsworth,N.J., Brehm,J.K., Atkinson,T.  
and Minton,N.P.

TITLE Complete nucleotide sequence of the Clostridium botulinum gene

encoding the type B neurotoxin

Unpublished (1991)

Location/Qualifiers

1..4041

/organism="Clostridium botulinum"

/db\_xref="taxon:1491"

57..3932

/gene="bont"

57..3932

/gene="bont"

/function="vertebrate neurotoxin"

/codon\_start=1

/transl\_table=1

/product="neurotoxin type B"

/protein\_id="AA02311.1"

/db\_xref="GI:144735"

/translation="WPVTNNFNNDPIDNNILMMPPFARGTGRYYKAFKITDRIM  
IPERYTFYKPEDNKSIGIFNROVCEYDYPDYLNTDKKNIFLTMTKLNRIKSK  
PIGEKLEMIINGIPYLGROVPLEFNTNIASTVTNKLISNPGEVRKKGIFANLII  
FGPGVPLENETIDIGIQNHFSREGFGIMQKFCVSVFNVEKNGKASIFNRR  
GYFSPALIMHVLHGLXIKVDLPIVNEKKFFMQSDAQABELTYFGQD  
PSIITPSTKSIYELKQVGRVGRVLRNKLVCISPDNININIKKFKVYFSDS  
ECKYSIDVESFQKLYKLMFGTETNAENYKIKTPASVSDSLPVLKLLDNEIY  
TEEGFNISDKMEYKQCNKAIKQAEIEISKEHLAVYIQMCKSVKAPGICIDVD  
NEDLFFADKNSFDDLSKNERIENYNTQSYIENDFPINELIDTDLISKIEPSENT  
ESLTFNVDVPIYKQPAIKKIFTENTFIQYLYSQTFPLDIRDLSLTSFDDALFS  
NRVYFFSDMYIKTANKVVEAGLFGAGVQIIVDFVEANKSNTMDKIADLSIVPI  
GLALNVNGTAKGNFENAFIAGASTILLEPILIPVVGAFLESYIDNNKIKITI  
DNALPKRNEKSDMYGLIYVAQLSVNTQFYTIKEGMYKALYQQAALBEIKRYNI  
YSEKSKNINIDFNLKNGEINQAIIDNININFGCSVSYLMKMKIPLAVEKLLDF  
DNLLKNNLYIDENKYLIGSAEYKSKVNYLKTIMPFDLSIYNTDILITEMFNKY  
NSEILLNINILNRYDNGVDSYKAKVEYDQVDELNDKQFKLTSANSKIRVTQN  
QNIIFNSDFSVSFWRIPKYKNDGIONYIHNEYTIICNMKNSGKISIRGNRII  
WTLIDNGKTSVFYREDISEVINRFEVFTINNNNAKVIYINGLESNTDKD  
IREVANGELIIFKLDGIDIRTQFIWKKYFSEIENFELSQSNIERYKIQSYSEYKDFW  
GNPLMTNRTYMFNAGKNSYIKLAKDSPVGLTFRSKYNKSKYINRYDLTIGERFI  
IRKNSNSINDIDYKQEDYIYDLFNLNOEWRVYTYFKKEEKLFLAPISDSDEF  
YNTIQIKDEQPTSCQLLKDEESTDEIGLTHREFYESGIVFEYKDYFCISKW  
YLKVEYKRPYNLKLGCMQWQIFPKDEGWT"

BASE COUNT 1679 a 383 c 645 g 1334 t

ORIGIN

alignment\_scores:

Quality: 2340.00 Length: 440

Ratio: 5.330 Gaps: 0

Percent Similarity: 99.773 Percent Identity: 99.545

alignment\_block:

US-09-910-186a-8 x CLOBOTB



```
/translation="MSVTINNFNNDPIDNDNIIMMEPPFARGTGGRYKAFKPIDRIW
IIPERYTEGYPEDENKSSGIFNRDVCEYIDPDYLTNDKKNFIQTMIKLFNRKSK
PLGEKLIEMINGIPYGLDRVRPLPEFNNTIASVTNKLISNPGVERKKGIFANLII
FGPGVPLNENETIDIGIONHFRASREGFGIMQKFCPEVSVFNNOENKGAIFNRR
CYFSPDALIHELHVLGLYGIKVNDLPVIVNEKKFFPMQSTDAIQAEELYTFGGQD
PSIISPDTKSIIYKVLQNFGRGIVDRLNKLVLCISDPNININIKNPKDKYKFVEDS
EGKYSIDVESDEKLYKSLMGFTETNIAENYKIKTRASYSDSLPPVKIKNLLDNEIY
TIEEGFNISDNMEKEYRGONKAIKQAYEIEISKEHLAVYIQMCKSVKAPGICIDVD
NEDLFIADKNSFSDLSKNERIAYTQNNYIDNDFSINELIDTDLISKTELPSNT
ESLTDENVVPEYKOPAKIKFTDENTIFOYLYSOTPLDIRDLSLTSSPDDALLFS
NKVYSFESMDYIKTANKVVEAGLFAGWVKQIVDDPVEIANKSSFWMDKADISLIVPYI
GLANVNGETAKGNFENAFETIAGASILLETIPPELLIPVYGFLLESYIDNKNKIETI
NSALTQRDEKIMDGLIIVLAQMLSTVQFYTIKEGMYKALNYQAOLEEIIKYKYNI
YSEKRSNINIDFNDVNSKLNENQAINDNINNEFNECSVSLMKMIMPLAVELKILDF
DNTLRKNLLINTIDENKLLIGSAEYKSKVDKIKLTSIPFOLSTVNTNTILIEIFNKY
NSDILLNILLNRYRDNKIDLSGVYKAVEYDVGVKLNKNOFKILTSSANSKIRVION
QNLIFNSMFLDPSVFWIRIPKYKNDGIONYIHNEYTLINCKNNSGKISIRGNMII
WTLIDINKIKSVFEYSIKEDISEYINRWFVTITNNSDKYITNKLKSHIDIRD
IREVITANDEIIFKLDGNIDRTQFIMKYFSIFNTELSQSNIEEIIKIQSYSEYKDFW
GNPLMYNKEYTFNAGNKSXYIKLKDSPVGEILTRSKYNQNSKYINRDLYIGSKFI
IRKNSQSQIINDDIIVRKEDYIYLDPFNLNQEWRYIYKYFKKEEKLFLAPISDSEF
YNTIQIKVDQPTVSCOLLFKKDEESTDEIGLIGHRFYESGIVFEKDYFCISKW
YLKVKRKPYNKGLGCKNQFIPKORGWTE"
```

BASE COUNT 1611 a 370 c 615 g 1280 t  
ORIGIN

alignment\_scores:      Quality: 2231.00      Length: 438  
                         Ratio: 5.200              Gaps: 0  
Percent Similarity: 97.945      Percent Identity: 94.977

alignment\_block:

US-09-910-186a-8 x AF295926

Align seg 1/1 to: AF295926 from: 1 to: 3876

```
3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTy 19
|||||
2560 AATAAATAATACCGATATTTAAATAATATATCTTAATTAAGATA 2609
|||||
19 rLysAspAsnLeuLeuAspLeuSerGlyTyrGlyAlaLysValGluV 36
|||||
2610 TAGGGATAATAAGTAAATAGATTATCAGGATATGGGGCAAGTAGAGG 2659
|||||
36 altYrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeuThrSer 52
|||||
2660 TATATGATGGGGTCAAGCTTAATGATAAAAAATCAATTTAAATTAACTAGT 2709
|||||
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
|||||
2710 TCAGCAATAGTAGAATAGAGTGATCAAAATCAGAAATATCATATTTAA 2759
|||||
69 nSerValPheLeuAspPheSerValSerPheThrIleArgIleProLys 86
|||||
2760 TAGTATGTCCTTCGATTTAGCGTTAGTTTTTGATAAAGAAATACCTAAAT 2809
|||||
86 yLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
|||||
2810 ATAGGAATGATGGTATACAAAAATATATTCATAATGAATACAAATAATT 2859
|||||
103 AsnCysMetLysAsnAsnSerGlyTyrLysIleSerIleArgGlyAsnAr 119
|||||
2860 AATTGATGAAAAAATATTCGATGGAAAAATATCTATTAGGGGTAATAT 2909
|||||
119 gIleIleThrPheLeuLeuAspIleAsnGlyLysThrLysSerValPheP 136
|||||
2910 GATAAATAGGACCTTAATTCATATAAATGAAAAATCAAAATCAGTATTT 2959
|||||
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrpPhe 152
|||||
2960 TTCAATATAGCATAAAGAAGATATATCAGAGTATATAAATAGATGGTTT 3009
|||||
```

seq\_name: gb\_ba:AF300465

```
153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnGI 169
|||||
3010 TTGTGTAACCTATTACTAATAATTCGGATAACGCATAAAATTTATATTATGG 3059
|||||
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAlaA 186
|||||
3060 TAAGCTAGAATCATATAGATATATTAGAGATATAAGAGAAGTTATTGCTA 3109
|||||
186 snGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|||||
3110 ATGATGAAATAAATAATTAAATTAGATGCTAATATAGATAGACACAAATTC 3159
|||||
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||
3160 ATTTGGATGAAATATTTTCAGTTATTTTAAATACGGAATTAAGTCAATCAA 3209
|||||
219 nIleGluGluArgTyrIlysIleGlnSerTyrSerGluTyrIleLysAsp 236
|||||
3210 TATTTGAAGAATAATATAAAATTTCAATCATATAGCGAATATTTAAAGATT 3259
|||||
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||
3260 TTTGGGGAAATCCTTTAATGTACATAAAGAATATATATATGTTTAATGCC 3309
|||||
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyGI 269
|||||
3310 GGGAATAAAAAATTCATATATTAACCTAAAGAAAGATTCACCTGTAGTGA 3359
|||||
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||
3360 AATTTTAAACAGTACGAGCAATATATCAAAATTTCCAAATATATAAATATA 3409
|||||
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSerAsnSer 302
|||||
3410 GAGATTTATATTTGGAGAAAAATTTATTATAGAGAAAGTCAAAATCTCT 3459
|||||
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319
|||||
3460 CAATCATATAATGATGATATAGTTAGAAAAGAGATTTATATATATCTAGA 3509
|||||
319 pPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysTyrPheL 336
|||||
3510 TTTTTTTAAATTTAAATCAAGAGTGGAGATATATATCTATAAAATATTTTA 3559
|||||
336 yLysGluGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
|||||
3560 AGAAAGAGGAAGAAAAATTTGTTTAGCTCCTATAAGTGATTTCTGATGAG 3609
|||||
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
|||||
3610 TTTTACAATACTATACAAATAAAGAATATGATCAACAGCAACATATAG 3659
|||||
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
|||||
3660 TTGTCAGTTGCTTTTAAAAAAGATGAAGAAGTACTGATGAGATAGGAT 3709
|||||
386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
|||||
3710 TGATGTGATTCATCGGTTTCTACGAATCTGGAATTTGATTAAAGAGTAT 3759
|||||
403 LysAspTyrPheCysIleSerLysTrpTyrLeuLysGluValLysArgLy 419
|||||
3760 AAAGATATTTTGTATAAGTAAATGCTACTTAAAGAGGTAAAAAGGAA 3809
|||||
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProLysAspG 436
|||||
3810 ACCATATAAATTCAAAATTTGGGATGTAATTTGGCAGTTTATTCTCTAAAGATG 3859
|||||
436 luGlyTrpThrGlu 440
|||||
3860 AAGGTGGGACTGAA 3873
```

```

seq_documentation_block:
LOCUS      AF300465          3876 bp    DNA        linear        BCT 08-OCT-2001
DEFINITION Clostridium botulinum isolate 588 type B cryptic neurotoxin gene,
            complete cds.
ACCESSION  AF300465
VERSION    AF300465.1  GI:15982936
KEYWORDS   .
SOURCE     Clostridium botulinum.
            Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
            Clostridium.
REFERENCE  1 (bases 1 to 3876)
AUTHORS   Kirma, N., Ferreira, J.L. and Baumstark, B.R.
TITLE     Characterization of six type A strains of Clostridium botulinum
            that contain type B toxin gene sequences
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 3876)
AUTHORS   Kirma, N., Ferreira, J.L. and Baumstark, B.R.
TITLE     Direct Submission
JOURNAL   Submitted (28-AUG-2000) Department of Biology, Georgia State
            University, P.O. Box 4010, Atlanta, GA 30302-4010, USA
ATTURES   location/Qualifiers
            1. .3876
            /organism="Clostridium botulinum"
            /isolate="588"
            /db_xref="taxon:1491"
            /country="USA: Ohio"
            /note="isolated from stool sample in 1976
            type: Ab"
            1. .3876
            /codon_start=1
            /transl_table=11
            /product="type B cryptic neurotoxin"
            /protein_id="AA11498.1"
            /db_xref="GI:15982937"
            /translation="MPVTINNFNYNDPIDNNIMMEPPFARGMGRRYKAFKFTIDRIW
            IIPERTFGYPDPDFNKGSGIFNRDCEYYDDPLATNDKKNIFLOTMIKLPNRKSK
            PLGEKLEMIINGPVLGDRRVPLEEFNTIASVTYVKNKLSNPFGEVERKKGIFANLII
            FGPGVLNETIDIGIONHFASREGFGGGMOKFCPEYVSFVNNQENKNGASTFNR
            GYSPDALILHBLIHVLHGLGIVKNDLPIVFNKKFFMQSDATAQAEELYTFGGQD
            PSIISSDKSIDYKVLQNFGRIVLRNKLVCISDPNININIKNKRDKYKFVEDS
            EGYSDIVSFDEKLYKSLMPGFTETNIAENYKIKTRASFSDSLPVKIKNLLDNIIY
            TIEGNSIDKMEKRYGQNKAINQAEIEISKEHLAVYKIQMCKSVKAPGICIDVD
            NEDLFTADKNSDDLSKNERIAYTONNYIENDFSINELILOTLSISFELPSENT
            ESITDFNVPVYVKQPAIKKIFTDENTIFQYLYSOTFPLDIRDISLTSFDDALLFS
            NKYVSFSDMYIKTANKVVEAGLFAGVKQIVNDFVIEANKSSTMBKDIADISLIVPI
            GLALNVGNETAKGFENAFETAGASILLFPELLIPVVGAFILLESIDNKNKIIETI
            NSALTNRKDEKIDMYGLIVAQWLSTVNTQFYTIKMGYKALNTYQAQALEEIKRYKNI
            YSEKRSNIINIDFVNSKLNQINQAINNFINECVSYLKMKMIPLAVEKLLDF
            DNTLRKNLLNYIDENKLYLGSAEYKSKVDYKLTSPDLSLTNTNTLIEFNKY
            NSDILANILNLRNKLIDLSYGAKVEVDGKLNKDNKFKLTSSANSKTRVTQN
            QNIFNSMFLDFSWSWIRIPKNDGIONYINNEYTIINCMKNNSGWLKISIRGNMII
            WTLIDNGKTSKVFESYKEDISEYINRFFVTITNNSDNKIIYINGKLESHDIRD
            IRVIANDEITFLDGNIDRTOPIWMKYFSIFNTELSQSNIEEYIQSYSEYKDFW
            GNPLMYNKYMFNAGKNYSIKLKDSSVGEILLTRSKYNSKSYINRYDLIGERKFI
            IRKNSQSNDDVIDKEDYIYLDENLNQEWRYMYKFKKEEKLFLAPISDSDEF
            YNTIQIKVEYDQPTYSQCLFKEKDEESTDEIGLIGTHRFVESGIVPEKDYFCISKW
            YLKEVRKPKNSKLGKNQWQFIPKDEGWTE"
BASE COUNT      1611 a      367 c      616 g      1282 t
ORIGIN

alignment_scores:
  Quality: 2229.00      Length: 438
  Ratio: 5.196          Gaps: 0
  Percent similarity: 97.945      Percent identity: 94.977

alignment_block:
US-09-910-186a-8 x AF300465
Align seg 1/1 to: AF300465 from: 1 to: 3876

```

```

3  AsnLysTyrAsnSerGluLeuAsnAsnIleIleLeuAsnLeuArgTy 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2560 AATAATAATAGCGATATTTAATAATATATATCTTAAATTTAAGATA 2609
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2610 TAGGGATAATAAGTTAATAGATTATCAGGATATGGGCAAGGAGTAGAG 2659
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
36 alTyrAspGlyValGluLeuAsnAspLysAsnGlnPheIysLeuThrSer 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2660 TATATGATGGGTCAAGCTTAATGATAAAATCAATTTAAATTAAGTACT 2709
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2710 TCAGCAATAGTAAGATTAGAGTGACTCAAAATCAGATATCATATTTAA 2759
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
69 nSerValPheLeuAspPheSerValSerPheTrpIleArgIleProLysT 86
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2760 TAGTATGTTCTTGTATTTTAGCGTTAGTTTGGGATAAGAATACCTAAAT 2809
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
86 rLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2810 ATAAGAATGATGGTATACAAATATATATTCATATGAATATACAAATAAT 2859
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
103 AsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAsnAr 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2860 AATTCTATCAAAATAATTTCTGGATGGAAATATCTATTAGGGGTAATAT 2909
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
119 gIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPheP 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2910 GATAATATGGACTTTAATTGATATATAATTCGAAAAATCAAAATCAGTATTT 2959
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrpPhe 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2960 TTGAATATAGCATAAAGAAGATATATCAGAGTATATAAATAGATGGTTT 3009
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnGl 169
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3010 TTTGTAACTATTACTAATAATTCGGATAACGCTAAAAATTTATATTAATGG 3059
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
169 yLysLeuGluSerAsnThrAspIleAspIleAspIleArgGluValIleAla 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3060 TAAGTAGAATCACATATAGATATTAGAGATATAAGAGAAGTATTGTCTA 3109
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
186 snGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3110 ATGATGAATAATAATTTAATAGATGGTAAATATAGATAGACACAGTTC 3159
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3160 ATTTGGATCAAAATATTTTCAAGTATTTTAAATACGGAATTAAGTCAATCAA 3209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAspP 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3210 TATTCAAGAATAATATAAAATTCATATACATATACGCAATATTTAAAAAGATT 3259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3260 TTTGGGAAATCCTTTAATGTACAATAAAGAATATATATGTTTAAATGCC 3309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyGl 269
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3310 GGGATAAAATTCATATATATAAACTAAAGAAAGATTTCATCTGTAGGTGA 3359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3360 AATTTTAAACACGTACCAATAATAATCAAAATTTCCAAATATATAAAATATA 3409
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSerAsnSer 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3410 GAGATTTATATTGGAGAAAAATTTATTATAGAAGAAAGTCAAAATCT 3459
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319

```



```
|||||.....:::|||||.....
3010 TTTGTAACATATTACTAATAAATTCGATACGCTAAATTTATATTAATGCG 3059
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAlaA 186
|||||.....:::|||||.....
3060 TAAGCTAGAAACACATATAGATATTAGATATAGAGAAAGTATTGCTA 3109
186 snGlyGluIleLeuPheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|| .....:::|||||.....
3110 ATGATGAATAATATTTAAATAGATGCTAATATAGATAGACACAGTTC 3159
203 lleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||.....:::|||||.....
3160 ATTTGGATGAATAATTTTCAGTATTTTAAATACGGAATTAAGTCAATCAA 3209
219 nileGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAsp 236
|||||.....:::|||||.....
3210 TATTGAAGAAATATATAAAATTCATCATATAGCGAATATTTAAAGATT 3259
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||.....:::|||||.....
3260 TTTGGGAAATCCTTAATGATACATAAAGAAATATATATGTTAATCGG 3309
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyG 269
|||||.....:::|||||.....
3310 GGGATAAAATTCATATATTAACATAAAGAAAGATTCATCTAGGTGA 3359
269 uileLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||.....:::|||||.....
3360 AATTTTAACACGTAGCAATATAATCAAAATTCCAATATATAAATATA 3409
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSerAsnSer 302
|||||.....:::|||||.....
3410 GAGATTTATATATTGGAGAAAAATTTATATAGAGAAAGATCAAAATCT 3459
303 GlnSerIleAsnAspAspIleValArgLysGluAspTyrIleTyrLeuAs 319
|||||.....:::|||||.....
3460 CAATCATATAATGATGATATAGTTAGAAAGAAAGATTTATATATCTAGA 3509
319 pPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysTyrPheL 336
|||||.....:::|||||.....
3510 TTTTTTTAAATTAATCAATCAAGAGTGGAGAGTATATATGATATAAATTTA 3559
336 ysLysGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
|||||.....:::|||||.....
3560 AGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3609
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
:::|||||.....:::|||||.....
3610 TTTTACAATATATACAAATAAAAGAAATATGATGAACAGCAACATATAG 3659
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
|||||.....:::|||||.....
3660 TTGTCAGTTGCTTTTAAAAAGAGTGAAGAAAGTACTGATGAGTAGAGT 3709
386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
|||||.....:::|||||.....
3710 TGATTTGTTATTCATCGTTCTACGAATCTGGAATGTTATTTAAAGAGTAT 3759
403 LysAspTyrPheCysIleSerLysTyrTrpTyrLeuLysGluValLysArgLy 419
|||||.....:::|||||.....
3760 AAAGATTATTTTGTATTAAGTAAATGGTACTTAAAGAGGTAAGAAAGGAA 3809
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProLysAspG 436
|||||.....:::|||||.....
3810 ACCATATATTAATCAAAATTTGGGATGTAATGGCAGTTTATTCCTCAAGATG 3859
436 luGlyTrpThrGlu 440
|||||.....:::|||||.....
3860 AAGGTTGGACTGAA 3873
seq_name: gb_ba:CBY13630
```

```
seq_documentation_block:
LOCUS      CBY13630                      11170 bp    DNA       linear       BCT 25-OCT-1998
DEFINITION      Clostridium botulinum HA-70, HA-17, HA-33, P-21, ntnh, bonf genes.
ACCESSION      Y13630
VERSION        Y13630.1
KEYWORDS        GI:3805779
               bonf gene; HA-17 gene; HA-33 gene; HA-70 gene; ntnh gene; P-21
               gene.
SOURCE          Clostridium botulinum.
ORGANISM        Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
               Clostridium.
REFERENCE      1 (bases 1 to 11170)
AUTHORS        Santos-Buelga,J.A., Collins,M.D. and East,A.K.
TITLE          Characterization of the genes encoding the botulinum neurotoxin
               complex in a strain of Clostridium botulinum producing type B and F
               neurotoxins
JOURNAL        Curr. Microbiol. 37 (5), 312-318 (1998)
MEDLINE        98440323
REFERENCE      2 (bases 1 to 11170)
AUTHORS        Santos-Buelga,J.A.
TITLE          Direct Submission
JOURNAL        Submitted (05-JUN-1997) J.A. Santos-Buelga, BBSRC Institute of Food
               Research, Earley Gate, Whiteknights Road, Reading, RG6 6BZ, UK
FEATURES       Location/Qualifiers
               1..11170
                /organism="Clostridium botulinum"
                /strain="CDC 3281 (ATCC 43757)"
                /db_xref="taxon:1491"
                /clone="PAET98-99"
                /clone="pJSB1-7"
                /clone="pJSB20"
                /clone="pJSB24"
                complement(1..1326)
                /gene="HA-70"
                complement(1340..1780)
                /gene="HA-17"
                complement(1340..1780)
                /gene="HA-33"
                complement(1340..1780)
                /codon_start=1
                /transl_table=11
                /protein_id="CAA73963.1"
                /db_xref="GI:3805780"
                /translation="MNSSIKKIYNDIQEKVINVSDTIDLADGNYVVRGQWILSRQN
                QILGVSINGSTGIVGLRVNDNAIPIYPTSPFNEEYIKNNIQVFTFNANQIP
                IGVEFSTAPSNKLYLYQIYIRYEIKVLQHEIIEIRAVLYVPSLGYKSTFEPNG
                EKINKDFYELTNDKILNQFLYKILETTKNPTNNIFNSKVSSTQRLVPLVSNGLYV
                INKDCYIRTNDDLLGLLEAGSSSIIQPRLNTTRPFTTSNDTKFSQOYTEER
                LDAPNVLNTSTSLKPEEAPSKNICIKAYYEKVELIDYQNGSVLNKAEYVL
                PSLGYCEVTNAPSPSESVKMQVAEDGFQNGPEEIVGVVIDPDSQENIQBINTAIDN
                YIYIPGIVNNPFIILFTVNTTGIYKINAQNNLPLPKIYEATGSGNRNN"
                complement(1340..1780)
                /gene="HA-17"
                complement(1340..1780)
                /gene="HA-33"
                complement(1340..1780)
                /codon_start=1
                /transl_table=11
                /protein_id="CAA73964.1"
                /db_xref="GI:3805781"
                /translation="MSVETFLPNGVNYNIKIFSGSLYNPVSLSLTFSSNANNOK
                WNVVEAENRCFKISNVAEPNKLYSDNFGFISLDSLSNRCWFPKIAVNTYIMLSL
                NKVNDIYAWDIYDTNENILSQPLLELPNFDIYNSQMFKEKI"
                complement(1843..2724)
                /gene="HA-33"
                complement(1843..2724)
                /gene="HA-33"
                complement(1843..2724)
                /codon_start=1
                /transl_table=11
                /protein_id="CAA73965.1"
                /db_xref="GI:3805782"
                /translation="MEHYSVIONSLNDKIVTISCRADTNLFFYQVAGNVSLSFOQTRNY
                LERWRIIYDSNKAAYKIKSDIHTNLVLTWNPNTIQTQDSNADNOYWLKLDIG
                NNSFIATSYKNPNLYADTVARNKLSTLNNSNYIKFIIEDYIISDFNNFTCKISPI
                LDNRKVVQOQVATNTLVNLYTWDYGRNOKWTIRYNEEKAAYQFFNTILNSGLTWLIFS
                NGNTVRVSSSDONNDNQWLNIPVSDTDETYITITNLRDITTKALDLYNSQTANGTAIQ
                VFNSSGDDNQKNIRNP"
```

gene  
CDS

```
2951..3487
/gene="p-21"
2951..3487
/gene="p-21"
/codon_start=1
/transl_table=11
/protein_id="CAA73966.1"
/db_xref="GI:3805783"
/translation="MNKLFQIEMLKNDNEEFQELFKHFEKTIINFTRKYNINYND
LYHLWLVKVDNSFNQNDLERYISRLKRYCLDICKNRKDKDKKLIYINSEADKK
LSLTANSYSSEFEFNDLISLIPDDQKKIYIMKFVEDIKEIDIATKLNISRSQSYKN
KIMALELEPIKLLNN"
3649..7242
/gene="ntnh"
3649..7242
/gene="ntnh"
/codon_start=1
/transl_table=11
/protein_id="CAA73967.1"
/db_xref="GI:3805784"
/translation="MNINDLSINSVDNKNVYVRBKTDTVFKAFKVAQINWVAPE
RYGESLSIDEXYKVDGIGYDSNLSQSKDKFLQALITLKRINSTNAGEKLSLI
STAIPFGYIGGGYAPNMITFGSAPKNKLSLISSTIPFPYAGYREKNTYSSSED
NKSYASINIFGPGCANIVENNTVYFKKDAENGMTWTEWFQFLLYKYDEYIDP
AIEIKLKLISLYFLYGIKPSDDLVIPIYKEDAEENGMTWTEWFQFLLYKYDEYIDP
NTDPWFDTNYSNAKKVFEDHNIYETEIEGNAIGNDIKLRLKOKFRININDIWEI
NLNYSKEFSIMPDRENNAKHFYRQYKIDYDPENYSINGFVNGQINAGLSDRN
QDIINKPREIINLNGNVSIMRSNIYDGLKSTVDYFISNKPYNRYEYHNNSN
DSSLDNVGIDNPEIIDNYPKENCDFSPQKISTREINTNIPWPINYLQAOQ
TNNEKSVSSDVEVVSQSKSLVYSLSFNMYFIDLSIKNSPIDTKKYYLWLREIF
RNYSFDTATQIINTDCINKVWTFGKALNLTSDSFVEEFQNLPSILNKREN
SMPIEYIENDMLGLPLNDLNEKLFNLYLKNIYFKKVFNFLDOWWTEVYSQYFD
LICMAQSILAQELIKOILQONKLDLPKADISMDKLNMLNATEKFFIDISNLSIA
LINNDLNLKSAICVFDNIPCKEISFEHQICINSVNSVTFPMQKCNITDEKQLLI
KLNFMNIDFFDIQSLKDLITSETDLIKEESDYNLFTLOEDNKNKVIDESGK
ITVLYSDISILSVGVGDALYLPKEDPESVSFNKAFENGLTNFSFSCFWLRNEDRI
ITSLKENADNCNGWEIYFENNGLVFSIVDCNGNEENIYLSDSIKNNWYISISDRL
NQULIFINDKLIQAQSEQLINLIYSNTISLVNENNPYIEGLSILNRSITSEVNN
NYFSLNNYSIRDTSGERLEYNKTYLVNYPENSLYEVTENNINILYSIKDTNNLI
QGAFFKLNIDANKQYQKWDGCVVCLLGDSEKYVDISSENNRQLVSSKDTAKRIIF
NDIFRPNCILTFATNNKLSLSLRDNTNMWICNNNNNNIPKAAHLWALKGI"
7268..11143
/gene="bont"
7268..11143
/gene="bont"
/codon_start=1
/transl_table=11
/protein_id="CAA73968.1"
/db_xref="GI:3805785"
/translation="MPVTINNFYNDPIDNNNIIMPEPPFARGMGRYVKAFKITDRIW
IIPERYFGYKPEDNFKSSGIFNRDCEYDYPDLNTNDKKNIFLOTWKILFNRIKSK
PLGEKLLMIINGIPYLGDRVPLEEFNTNLTASVTNKNLSNPGVEVERKKGIPANLII
FGPGVLNENTIDIGIQNHPASREGFGIGMQMPCPEYVSFVNNVQENKASIFNRR
GYFSDPALIMHELIHLVHLYGIVKNDLPVNPKEKFFMQSDAIQAEELITFGGDQ
PSIISPSDKSIYDKVQNFRIQVDRNLKVLVCSIDPNININIKNFKPKRYKFEVDS
BGKYSIDVESDKLYKSLMFQGTETNIAENYIKTRAAYFSDSLPPVKPKNLLDNEIY
TIEGFFNSDKANMEKEYGONKAIKQAYEISKEHLAVYKIOMCKSVKAPGICLDVD
NEDLFTADKNSFSDLSKNERIAYNTQNNYIENDFSINELILOTDLISKIELSENT
ESLIDFNVPYVKQPAIKKIFTIDENTIFQIYLSQTFPLDIRLSLFSSEFDALLFS
NKVYSFSDMYIKTANKVVEAGFAGWKQIVDDFVIBANKSSYMDKIDLISLVPIYI
GALNNGNETAKGFNAFETAGSILLETPELLIPVVGAFLESYIDNKNKIETI
NSALTQRDEKVIDMGLYVAQLSVTNQFTYIKEGMYKALNYOAOALEEIKYKNI
YSERKSNINIDFNDVNSKLNQINQADINNNFINECVSYLMMKMTPLAVEKLDF
DNTLRKLLNIDENKLYLIGSAEYKSKVKYKLTSTIPDLSTYNTNLTILFENKY
NSDILNLIILNRYDRNDKLDLSYGAKVEYDVGKLDNRKNQFLTSSANSKIRVION
QNIIFNSMDFSVSWFIRIPKNDGQNTYIHNEYTIINCKMNSGWKISIRGNMII
WTLIDINGIKSVFEFYSIKEDISEYINRFEVFTTNNSDNKIYINKGLSHDIDRD
IREVTADEILFLKDGNDTRQFTWMTKYFSENFELSOSNEEYIKIOSYSEYLDQFW
GNPLMYNKEYYFNAGKNKNSYIKLKQSSVGEIILTRSKYNONSKEYINRDLIIGEKFI
IRKNSQSINDDIIVRKEDYIYLDFFNLQNEQWRYMYKFKKEBEKFLAPDISDEF
YNTIOIKEDQPTYSQQLLFPKDEESTDETLGIIHRFYESIVFKEYKDFYCISKW
YLKVEVKRPYNSKLGCNQFIPKDEGWE"
```

gene  
CDS

```
3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTy 19
|||||
9827 AATAAATAATAACGCTTAATTAATAATAATATATCTTAAATTTAAGATA 9876

19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36
|||||
9877 TAGGGATAATAAGCTTAATAGATTTATCAGCATATGGGGCAAGGTAGAGG 9926

36 alTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeuThrSer 52
|||||
9927 TATATGATGGGGTCAAGCTTAATGATAAAAAATCAATTTAAATTAACCTAGT 9976

53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
|||||
9977 TCAGCAAAATAGTAAGTATAGAGTGATTCAAAATCAGAAATATCATATATA 10026

69 nSerValPheLeuAspPheSerValSerPheTrpIleArgIleProLysT 86
|||||
10027 TACTATGTCTCTTGATTTTAGCGTGTAGTTTGGATAAAGAAATACCTAAAT 10076

86 yLysAsnAspGlyIleGlnAsnTrpIleHisAsnGluTyrThrIleIle 102
|||||
10077 ATAAGAATGATGTGTATACAAAAATATATTCATAATGAATATACAATAATT 10126

103 AspCysMetLysAsnAsnSerGlyTyrLysIleSerIleArgGlyAsnAr 119
|||||
10127 AATTGTATGAAAAATTAATCTCGATGGAAAAATATCTATTAGGGGTAATAT 10176

119 gIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPheP 136
|||||
10177 GATAATATGGACTTTAATTCATATAAATGCAAAATCAAAATCAGTATTTT 10226

136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrpPhe 152
|||||
10227 TTGAATATAGCATAAAAGAAAGATATATCAGAGTATATAAATAGATGTTT 10276

153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnGl 169
|||||
10277 TTTGTAACTATTACTAATAATAATTCGGATAACGCTAAAATTTATTAATGG 10326

169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAlaA 186
|||||
10327 TAAGCTAGAAATCATATACATATATAGAGATATAAAGAGAAGTTATTGCTA 10376

186 snGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|||||
10377 ATGATGAAATAAATATTAAATTTAGATGTTAGTAAATATAGATAAGACACAGTTC 10426

203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||
10427 ATTTGGATGAAATATTTCAGTATTTTAAATACGAAATTAAGTCAATCAAA 10476

219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAsp 236
|||||
10477 TATTGAAGAAATATAAATAATTCATCATATACGGAATATTTAAAGCATT 10526

236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||
10527 TTTGGGGAATCCCTTTAAATGTAACAATAAAGAAATATATATATGTTTAAATGCG 10576
```

BASE COUNT

4359 a 1172 c 1642 g 3997 t



```

253 GlyAsnLysAsnSerTyrIleLeuLysLeuLysLysAspSerProValGlyCl 269
10577 GGGATAAAATTCATATATTAACTAAAGAAAGATTTCATCTGAGGTCA 10626
269 uileLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
10627 AATTTTACAGCTGCAAAATATAATCAAAATCCAAATATATAAAATTATA 10676
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgLysSerAsnSer 302
10677 GAGATTATATATTGGAGAAAATTTATATAGAAGAAAGATCAAAATCT 10726
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319
10727 CAATCTATAAATGATGATATAGTTAGAAAGAAAGATATATATATCTAGA 10776
319 pPhePheAsnLeuAsnGlnGluTyrArgValTyrThrTyrLysTyrPheL 336
10777 TTTTTTTTAAATCAAGAGTGGAGAGATATATATGATATAAATATTTTA 10826
336 ysLysGluGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
10827 AGAAAGAGAGAAATAATGTTTGTAGTCTCTATTAAGTGTCTGATGAG 10876
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
10877 TTTTCACTACTATACAAATAAAAGAAATATGATGAACAGCCCAACATATAG 10926
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
10927 TTGTCAGTTGCTTTTAAAGAGATGAAGAAAGTACTGATGAGATPAGGAT 10976
386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
10977 TCAATGGTATTCATCGTTCTACGAATCTGGAATGTGATTTAAAGAGTAT 11026
403 LysAspTyrPheCysIleSerLysTrpTyrLeuLysGluValLysArgLys 419
11027 AAAGATTATTTTGTATTAAGTAAATGGTACTTAAAGAGGTAAAGAGGAA 11076
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProLysAspG 436
11077 ACCATATAATCAAAATTTGGATGATTAATGGCAGTTTATCTCTAAAGATG 11126
436 luGlyTrpThrGlu 440
11127 AAGGTGCTGACTGAA 11140

```

seq\_name: gb\_ba:AF300467

seq\_documentation\_block:

CUS AF300467 3869 bp DNA linear BCT 08-OCT-2001  
 DEFINITION Clostridium botulinum isolate 519 type B cryptic neurotoxin-like  
 gene, complete sequence.

ACCESSION AF300467

VERSION AF300467.1 GI:15982940

KEYWORDS Clostridium botulinum.

SOURCE Clostridium botulinum

ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

Clostridium.

REFERENCE 1 (bases 1 to 3869)

AUTHORS Kirma,N., Ferreira,J.I. and Baumstark,B.R.

TITLE Characterization of six type A strains of Clostridium botulinum

that contain type B toxin gene sequences

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3869)

AUTHORS Kirma,N., Ferreira,J.I. and Baumstark,B.R.

TITLE Direct Submission

JOURNAL Submitted (28-AUG-2000) Department of Biology, Georgia State

University, P.O. Box 4010, Atlanta, GA 30302-4010, USA

COMMENT NCBI staff are still waiting for submitters to provide appropriate

coding region information.

FEATURES  
 source  
 Location/Qualifiers  
 1. 3869  
 /organism="Clostridium botulinum"  
 /isolate="519"  
 /db\_xref="taxon:1491"  
 /country="USA: Alaska"  
 /note="isolated from stool sample in 1976"  
 type: A(B)  
 misc\_feature  
 1. 3869  
 /note="similar to type B cryptic neurotoxin"  
 BASE COUNT 1607 a 365 c 614 g 1283 t  
 ORIGIN

alignment\_scores:  
 Quality: 2210.00 Length: 438  
 Ratio: 5.140 Gaps: 0  
 Percent Similarity: 98.174 Percent Identity: 94.521  
 alignment\_block:  
 US-09-910-186a-8 x AF300467 ..  
 Align seg 1/1 to: AF300467 from: 1 to: 3869

```

3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTyr 19
2554 AATAAATATAATAGCGATATTTTAAATAATATATATCTTTAAATTTAAGATA 2603
19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36
2604 TAGGATATAAAGTTAATAGATTTTATCAGGATATGGGCAAGGTAGAGG 2653
36 alTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeuThrSer 52
2654 TATATGATGGGTCGAGCTTAATGATAAAATCAATTTAAATTAAGTACT 2703
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
2704 TCAGCAATAGTAGATAGAGTACTCAAAATCAGAAATATCATATTTAA 2753
69 nSerValPheLeuAspPheSerValSerPheTrpIleArgIleProLysT 86
2754 TAGTATGTTCTTGAATTTAGCGTTAGTTTTTGGATAAGAAATACCTAAAT 2803
86 yrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
2804 ATAAGAATGATGGTATACAAAATATATATTCATATGAATATACATAATT 2853
103 AsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAsnAr 119
2854 AATTGTATGAAAAATAATCTCGGATGGAATA.TCTATTAGGGGTAATAT 2902
119 gIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPheP 136
2903 GATATATGGACTTTAATTGATATAAATGGAAAAATCAAAATCAGTATTTT 2952
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrpPhe 152
2953 TTGAATATAGCATAAAACAAGATATATCAGAGTATATAAATAGATGGTTT 3002
153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnGl 169
3003 TTTGTAACTATTACTATAATTCGCGATACCGTAAATTTATATTAATGG 3052
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAlaA 186
3053 TAAGCTAAGATCACATATAGATATATTAGAGATATAAGACAGATTATTGCTA 3102
186 snGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
3103 ATGATGAAATAATATTTAAATTAGATGCTAATATATAGATAGAACACAATTC 3152
203 lleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219

```



```
|||||
3153 ATTGGATGAATATTTCAGTGTGTTTTTAATACGGAATTAAGTCAATCAA 3202
|||||
219 nileGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysasp 236
|||||
3203 TATTGAAGAAGATATAAAATTCATCATATAGCGAATATTTAAAGATT 3252
|||||
236 heTPrGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||
3253 TTGGGGAAATCCTTAAATGACAAATAAGAAATATATATGTTAAATGCG 3302
|||||
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyG 269
|||||
3303 GGGAATAAAATTCATATATTAACATAAGAAAGATTCCACCTGTAGGTGA 3352
|||||
269 uileLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||
3353 AATTTTAACACGAGCAATATATATCAAAATTCCTCAATATATAAATATA 3402
|||||
286 rGAspLeuTyrIleGlyGluLysPheIleArgArgLysSerAsnSer 302
|||||
3403 GAGATTTATATATGGAGAAAATTTATATATAGAGAAAGTCAAAATCT 3452
|||||
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319
|||||
3453 CAATCTATAAATGATGATATAGTTAGAAAAGAGATTATATATATCTAGA 3502
|||||
319 pPhePheAsnLeuAsnGlnGluTyrPargValTyrThrTyrLysTyrPhe 336
|||||
3503 TTTTTTTAAATTTAAATCAAGAGTGGAGATATATATCTATAAATATTTTA 3552
|||||
336 yLysGluGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
|||||
3553 AGAAGAGGAGAGAAAATTTGTTTACCTCTATAGTGAATCTGATGAT 3602
|||||
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
|||||
3603 TTTTACAATACTATACAAATAAAGAAATATGATGAACAGCAACATATAG 3652
|||||
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
|||||
3653 TTGTCAGATTCCTTTTAAAGAAATGAAGAAAGTACTGATGATAGGAT 3702
|||||
386 eulleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
|||||
3703 TGATGGTATCATCGTTCTACGAATCTGGAATGTAATTTAAAGAGAT 3752
|||||
403 LysAspTyrPheCysIleSerLysTyrTyrLeuLysGluValLysArgLy 419
|||||
3753 AAAGATTATTTTGATAGCTAAATGCTACTTAAAGAGGTAAAGAGAA 3802
|||||
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProLysAspG 436
|||||
3803 ACCATATAATTCAAAATTTGGGATGTAATTGGAAGTTTATTCCTAAAGATG 3852
|||||
436 luGlyTyrThrGlu 440
|||||
3853 AAGGGTGGACTGAA 3866

seq_name: gb_ba:AF300468

seq_documentation_block:
LOCUS AF300468
DEFINITION Clostridium botulinum isolate 667 type B cryptic neurotoxin-like
gene, complete sequence.
ACCESSION AF300468
VERSION AF300468
KEYWORDS AF300468.1 GI:15982941
SOURCE Clostridium botulinum.
ORGANISM Clostridium botulinum
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 3869)
```

```
AUTHORS Kirma,N., Ferreira,J.L. and Baumstark,B.R.
TITLE Characterization of six type A strains of Clostridium botulinum
that contain type B toxin gene sequences
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3869)
AUTHORS Kirma,N., Ferreira,J.L. and Baumstark,B.R.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-2000) Department of Biology, Georgia State
University, P.O. Box 4010, Atlanta, GA 30302-4010, USA
COMMENT NCBI staff are still waiting for submitters to provide appropriate
coding region information.
FEATURES
source
1..3869
/organism="Clostridium botulinum"
/isolate="667"
/db_xref="taxon:1491"
/country="USA: Wisconsin"
/note="Isolated from stool sample in 1976
type: A(B)"
misc_feature
1..3869
/note="similar to type B cryptic neurotoxin"
BASE COUNT 1607 a 365 c 614 g 1283 t
ORIGIN
alignment_scores:
Quality: 2210.00 Length: 438
Ratio: 5.140 Gaps: 0
Percent Similarity: 98.174 Percent Identity: 94.521
alignment_block:
US-09-910-186A-8 x AF300468 ..
Align seg 1/1 to: AF300468 from: 1 to: 3869
3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTy 19
|||||
2554 AATAAATAATAATAGCGATATTTAAATAAATATATATCTTAAATTAAGATA 2603
|||||
19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36
|||||
2604 TAGGGATAATAAGTTAATAGATTTATCAGGATATGGGGCAAGGTAGAGG 2653
|||||
36 alTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeuThrSer 52
|||||
2654 TATATGATGGGGTCAGACTTAATGATAAAATCAATTTAAATTAAC TAGT 2703
|||||
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
|||||
2704 TCAGCAATAAGTAGAAGATTACAGTGACTCAAAATCAGAATATCATATTTAA 2753
|||||
69 nSerValPheLeuAspPheSerValSerPheTyrPheIleArgIleProLys 86
|||||
2754 TAGTATGTTCCCTTGATTTTAGCGTTAGTTTTTGGATAAAGAAATACCTAAAT 2803
|||||
86 yLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
|||||
2804 ATAAGAAATGATGGTATACAAAATTTATATTCATATGAATATACAAATAAT 2853
|||||
103 AsnCysMetLysAsnAsnSerGlyTyrLysIleSerIleArgGlyAsnAr 119
|||||
2854 AATTGATAGAAAATAATTTCTGGATGGAATA.TCTATTAGGGGTAATAT 2902
|||||
119 gIleIleTyrThrLeuIleAspIleAsnGlyLysThrLysSerValPheP 136
|||||
2903 GATAATATGGACTTTAATTCATATATAAATGAAAAATCAAAATCAGTATTTT 2952
|||||
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTyrPhe 152
|||||
2953 TTGAATATAGCAATAAAGAGATATATCAGAGTATATAAATAGATGTTT 3002
|||||
153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnGl 169
|||||
```

```
3003 TTTGTAACCTATTACTAAATAAATTCGGATACGCTAAATAATTTATATTAATGG 3052
169 yLysLeuGluSerAsnThrAspIleLeuAspIleArgGluValIleAlaAa 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3053 TAAGCTAGATACATATAGATATATAGAGATATAGAGAGTATTGCTA 3102
186 snglycyllelelphelyleuaspglyaspIleaspargthrGlnPhe 202
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3103 ATGATGAATAATATTTAAATTAGATGGTAAATAGATAGAACACAATTC 3152
203 iletrpmetLysTyrPheSerilePheAsnThrGluLeuSerGlnSerAs 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3153 ATTTGGTGAATATTTTCAGTCTTTTAAATACGGAATTAAGCAATCAAAA 3202
219 nileGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAspP 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3203 TATTGAAGAAAGATATAAATTCATCATATATAGCGAATATTTAAAGATT 3252
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3253 TTTGGGAAATCCTTTAATGTACAATAAAGATATATATATGTTAATGCG 3302
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyCl 269
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3303 GGGAAATAAAATTCATATATTAACATAAAGAAAGATTCACCTGTAGGTGA 3352
269 utileLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3353 AATTTTACACGTAGCAATATATCAAAATTCCAAATATATATAATATA 3402
286 rgAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSerAsnSer 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3403 GAGATTATATATTTGGAGAAAAATTTATATAAGAAAGAAAGTCAAAATTC 3452
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3453 CAATCTATAAATGATGATAGTATAGAAAGAAAGATATATATATCTAGA 3502
319 pPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysTyrPheL 336
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3503 TTTTTTTAAATTAATCAAGAGTGGAGAGTATATATCTATAAATATTTA 3552
336 ySLysGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3553 AGAAAGAGAGAAAGAAATTTGTTTATAGCTCTATAAGTGATTCGTGATG 3602
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3603 TTTTACATATCTATACAAATAAAGAAATATGATGACACCAACATATAG 3652
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3653 TTGTCAATGCTCTTTTAAAAAGATGAAGAAAGTACTGATGATAGGAT 3702
386 euileGlyIleHisargPheTyrGluSerGlyIleValPheGluGluTyr 402
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3703 TGATTGGTATTCATCGTTCTACGAATCTGGAATTCGATTTAAAGAGTAT 3752
403 LysAspTyrPheCysIleSerLysTrpTyrLeuLysGluValLysArgLy 419
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3753 AAAGATATTTTGTATAGTAATATGGTACTTTAAAGAGGTAAAGAGGAA 3802
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProLysAspG 436
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3803 ACCATATAATTCAAAATTTGGGATGTAATTTGGAAGTATTATTCCTAAAGATG 3852
436 luGlyTrpThrGlu 440
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3853 AAGGGTGGACTGAA 3866
seq_name: gb_ba:AF300469
seq_documentation_block:
```

```
LOCUS AF300469 3869 bp DNA linear BCT 08-OCT-2001
DEFINITION Clostridium botulinum isolate 13280 type B cryptic neurotoxin-like
gene, complete sequence.
ACCESSION AF300469
VERSION AF300469.1 GI:15982942
KEYWORDS Clostridium botulinum.
SOURCE Clostridium botulinum
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 3869)
AUTHORS Kirma,N., Ferreira,J.L. and Baumstark,B.R.
TITLE Characterization of six type A strains of Clostridium botulinum
that contain type B toxin gene sequences
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 3869)
TITLE Direct Submission
JOURNAL Kirma,N., Ferreira,J.L. and Baumstark,B.R.
COMMENT Submitted (28-AUG-2000) Department of Biology, Georgia State
University, P.O. Box 4010, Atlanta, GA 30302-4010, USA
NCBI staff are still waiting for submitters to provide appropriate
coding region information.
FEATURES
source Location/Qualifiers
1..3869
/organism="Clostridium botulinum"
/isolate="13280"
/db_xref="taxon:1491"
/country="USA: Colorado"
/notes="isolated from peppers in 1972
type: A(B)"
misc_feature 1..3869
/notes="similar to type B cryptic neurotoxin"
BASE COUNT 1607 a 365 c 614 g 1283 t
ORIGIN
alignment_scores:
Quality: 2210.00 Length: 438
Ratio: 5.140 Gaps: 0
Percent Similarity: 98.174 Percent Identity: 94.521
alignment_block:
US-09-910-186a-8 x AF300469 ..
Align seg 1/1 to: AF300469 from: 1 to: 3869
3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTyr 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2554 AATAAATATATAGCGATATTTTAAATATATATATCTTAAATTTAAGATA 2603
19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36
|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2604 TAGGATAATAAGTTAATAGATTTATCAGGATATGGGCAAGGTAGAGG 2653
36 alTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeuThrSer 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2654 TATATGATGGGTCAAGCTTAATGATAAAATCAATTTAAATTAACACTAG 2703
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2704 TCAGCAATAGTAAGATTAGAGTGACTCAAAATCAGATATCATATTTAA 2753
69 nSerValPheLeuAspPheSerValSerPheTrpIleArgIleProLysF 86
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2754 TAGTATGTTCCCTGATTTTAGCGTTAGTTTGGATAAGAAATACCTAAAT 2803
86 yrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2804 ATAAGAATGATGGTATACAAAATTTATTCATAATGAATATACATAAAT 2853
103 AsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAsnAr 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2854 AATTGTATGAAAAATAATTCTGGATGGAAATA.TCTATTAGGGGTAAATAT 2902
```

```
119 gillelletpThrLeuileAspIleAsnGlyLysThrLysSerValPheP 136
|||||
2903 GATAAATGGACITTAATTCATATAAATGGAATAATCAAAATCAGATATTT 2952
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrrPhe 152
|||||
2953 TTGAATATAGCATAAAAGAGATATATCAGAGTATATAAATAGATGCTTT 3002
153 PheValThrIleThrAsnAsnLeuAsnAlaLysIleTyrIleAsnG1 169
|||||
3003 TTGTAACTATTACTAATAATTCGGATAACGCTAAAAATTTATATTAATGG 3052
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAlaA 186
|||||
3053 TAAGCTAGATCACAATATAGATATAGAGATATAAGAGAGATTATTCCTA 3102
186 snGlyGluIlePheGlyLeuAspGlyAspIleAspArgThrGlnPhe 202
|||||
3103 ATGATGAATAAATATTTAAATTAGATGCTAATATAGATAGACACAATTC 3152
203 IleTrrPheMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||
3153 ATTTGGATGAATAATTTCACTGCTTTTAAATACGGAATTAAGTCAATCAAA 3202
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAspP 236
|||||
3203 TATTGAAGAAGATATAAAATTTCAATCATATACGGAATATTTAAAGATT 3252
236 heTrrPlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||
3253 TTTGGGAAATCCTTTAATGTACAATAAAGAATATATATGTTAATGCG 3302
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyG1 269
|||||
3303 GGGAAATAAAATTCATATATAAATCAAAAGAAAGATTCACTGTAGGTGA 3352
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||
3353 AATTTAACACGTAGCAAAATATATCAAAATTCAAATATATATAATATATA 3402
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgLysSerAsnSer 302
|||||
3403 GAGATTTATATATGGAGAAAAATTTATATAAGAGAAAGTCAAAATCT 3452
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319
|||||
3453 CAATCTATAAATGATGATATAGTTAGAAAAGAGATTATATATATCTAGA 3502
319 pPhePheAsnLeuAsnGlnGluTrrArgValTyrThrTyrLysTyrPheL 336
|||||
3503 TTTTATTAATTTAAATCAAGAGTGGAGAGTATATATCTATAAATATTTTA 3552
336 yLysGluGluGluLysLysPheLeuAlaProIleSerAspSerAspGlu 352
|||||
3553 AGAAAGAGGAAGAAAAATCTTTTAGCTCCTATAGTATGATTTCTGATGAT 3602
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
|||||
3603 TTTTACAATACTACTACAAATTAAGAAGATATGATGAACAGCAACATATAG 3652
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
|||||
3653 TTGTAGTGTGCTTTTAAAAAAGATGAAGAAGTACTGATGAGATAGGAT 3702
386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
|||||
3703 TGATTTGATTCATCGTTTCTACGAATCTGGAATTTGATTTAAAGAGTAT 3752
403 LysAspTyrPheCysIleSerLysTrrPyrTyrLeuLysGluValLysArgL 419
|||||
3753 AAAGATTTATTTTGTATAAGTAAATGCTACTTAAAAAGAGCTAAAAAGGAA 3802
```

```
419 sProTyrAsnLeuLysLeuGlyCysAsnTrrPInPheIleProLysAspG 436
|||||
3803 ACCATAAATTCAAAATTTGGATGTAATTTGGAAGTTTATTCCTTAAAGATG 3852
436 luGlyTrrPThrGlu 440
|||||
3853 AAGGTTGGACTGAA 3866
seq_name: gb_ba.CBBONTB
seq_documentation_block:
LOCUS CBBONTB 4051 bp DNA linear BCT 20-APR-1994
DEFINITION C.botulinum type B gene for neurotoxin.
ACCESSION X71343
VERSION X71343.1 GI:296148
KEYWORDS botNT/B gene; botulinum neurotoxin type B; neurotoxin type B.
SOURCE Clostridium botulinum.
ORGANISM Clostridium botulinum.
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE
1 (bases 1 to 4051)
Hutson,R.A., Collins,M.D., East,A.K. and Thompson,D.E.
Nucleotide sequence of the gene coding for non-proteolytic
Clostridium botulinum type B neurotoxin: comparison with other
clostridial neurotoxins
Curr. Microbiol. 28 (2), 101-110 (1994)
94122659
REFERENCE
2 (bases 1 to 4051)
Hutson,R.A.
Direct Submission
Submitted (06-APR-1993) R.A. Hutson, AFRC Institute of Food
Research, Reading Laboratory, Microbiology Dept., Earley Gate,
Whiteknights Road, Reading, RG6 2EF, UK
Location/Qualifiers
1..4051
/organism="Clostridium botulinum"
/strain="Eklund 17b ATCC25765"
/isolate="type B"
/db_xref="taxon:1491"
/clone="pCBOTB1-8"
83..91
97..3972
/gene="botNT/B"
97..3972
/gene="botNT/B"
/codon_start=1
/transl_table=11
/product="BotNT/B"
/protein_id="CAA50482.1"
/db_xref="GI:296149"
/db_xref="SPTREMBL:Q08077"
/translation="MPVTINNFYNDPIDNDNIIMHPEFPARGTGRTGYKAFKITDRIV
IIPERYTFYKPEDNFKSGIFNDVCEYIDPDYLTNDKKNIFPOTLIKLFNRIKSK
PLGKLEMIINGIPYLGDRVRPVEEFNTIASVTNKLISNPGVEVRKKGIFANLI
FGPGVLNENETIDIGIONHFASREGFGIMQMKFCPEVSVSVNVOENKGIENRR
GYSPDALILMHELJHVLHGLYGIKVDOLPVPNEKKFFMOSTDITQAEELTFGQSD
PSLISPTDKSIYDKVLQNFGRVDRNLKVLVCISDPNINIYNNKFKDKYKFVEDS
EGKYSIDVESFNKLYKSLMGLFTEINIAENYKIKTRASTYSDSLPPVKIKNLDNEIY
TIEEFNISDKNMKEYGQNKAIKQAYEIEISKEHLAVYIQMCKSVKVPVICIDVD
NENLFFIADKNSFSDLSKNERVEYNTQNNYIGNDFPINELITDLISKIELSENT
ESLTDNFVDVPVYEKOPAIKKVFTDENTIFOYLYSOTFPLNTRDISLTSSPFDALLVS
SKVYSFFSMVDYIKTANKVEAGLFAGVWKQIVDDFVIEANKSSMTDKLADISLIVPYI
GLALNVGDETAKGNFESAFETAGSSILFELIPVVGFFLLESYIDNKNKIIRTI
DNALTRKVEKWIMDYG1VAQMLSTVTFYTIKEGMKALNYQAQALEEILKYKINI
YSEEKSNININFDNISKLANDINQAMNDINDFINECSVSLMKMIPLAVKLLDF
DNTLKNLLNIDENKLYIGSVEDEKSKVDKLYTIIPFDLSYNSIETLIKFNKY
NSEILNNTILNLRYDRNLDLSGYAKVEVDGVKLNQKOFKITSSADSKIRVTQN
QNILFNSMFLDSVSWFRIPIKYRNDLQNIHREYTIINCMMKNSCHKISIRGNRII
WTLDINGTKSVFFYREIDISEYINRWFVTITNLDNAKITINGTLESNDMDIKD
IGEVTVNGEITFKLGDVDYTFQIMWKYSIFNTQLNQSNIKETIKIOSYSEYLDKF
GNPLMYNKEYYFNAGKNKSYIKLVKDSVSGEILIRSKYNQNSNINIRNYLIGSKFI
IRRESNQSINDDIVRKEDYIHLDLVLHHEWRVYAYKYFKDEEKLFLSIISDSNEF
YKTIIEKEYDQPSYSCOLLFKKDEESTDGLIGLHRYFESGVLRKKYKDYFCISKW
```

YLKVRKKPKYKSLGCMQWFPKDEGWTE"

BASE COUNT 1679 a 384 c 645 g 1343 t  
ORIGIN

## alignment\_scores:

Quality: 2099.00 Length: 438  
Ratio: 4.962 Gaps: 0  
Percent Similarity: 96.575 Percent Identity: 88.813

## alignment\_block:

US-09-910-186a-8 x CBBONTB ..

Align seg 1/1 to: CBBONTB from: 1 to: 4051

3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTy 19  
|||||  
2656 AATAAATAAATAGCGAAATTTTAAATAATATATCTTTAAATTTAAGATA 2705  
19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36  
|:::|||||  
2706 TAGAGATAATAATTTAATAGATTATCAGGATATGGAGCAAGGTAGAGG 2755  
36 alTyraSpGlyValGluLeuAsnAspLysAsnGlnPheLysLeuThrSer 52  
|||||  
2756 TATATGATGGGTCAGGCTTAATGATATAAAATCAATTTAAATTAAGT 2805  
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69  
|||||  
2806 TCAGCGATAGTAAGATTAGAGTCACCTCAAAATCAGAATGATATATATTA 2855  
69 nSerValPheLeuAspPheSerValSerPheTrpIleArgIleProLysT 86  
|||||  
2856 TAGTATGTTCCCTTGAATTTAGCGTTAGCTTTTGGATAAGGATACCTAA 2905  
86 yLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102  
|||||  
2906 ATAGGAATGATGATATACAAAATTTATTTCAATAATGAATATACGATA 2955  
103 AsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAsnAr 119  
|||||  
2956 AATTGATGCAAAAATTAATTCAGCGTGGAAAATATCTATTAGGGGTAAT 3005  
119 gLileIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPheP 136  
|||||  
3006 GATAATATGGACCTTAATTGATATAAATGGAAAAACCAATCAGTATTT 3055  
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrpPhe 152  
|||||  
3056 TTCAATATATAACATAGAGAAGATATATCAGAGTATATAAATAGATGTT 3105  
153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnGl 169  
|||||  
3106 TTGTGAACATTACTAATAATTTGGTAATGCTAAATTTATATTAATG 3155  
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAlaA 186  
|:::|||||  
3156 CAGGTAGAAATCAAAATATGATATTAAGATATAGGAGAAGTTATTGTT 3205  
186 snGlyGluIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202  
|||||  
3206 ATGGTGAATAACATTTAAATTAGATGGTGGTATAGATAGAACACAA 3255  
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219  
|||||  
3256 ATTGGATGAATATTTTATGATTTTAAACCAATTAACCAATTAACATCA 3305  
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAsp 236  
|||||  
3306 TATTAAGAGATATAAAATTAATCAATCATATAGCGAATACCTTAAAG 3355  
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252  
|||||

3356 TTTGGGAAATCCTTTAATGTATAATAAAGAATATATATATGTTTATGCG 3405  
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyGl 269  
|||||  
3406 GGAATAAATAATTCATATATTAAGTAACTAGTGAAGAGATTCATCTGTAG 3455  
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286  
|||||  
3456 AATATTAATACGTAGCAAAATATAATCAGAAATCCAAATATATAAATATA 3505  
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSerAsnSer 302  
|:::|||||  
3506 GAAATTTATATATGGAGAAAATTTTATATAAGAGAGAGCAAAATCT 3555  
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrIleLeuAs 319  
|||||  
3556 CAATCTAATAATGATGATATAGTTAGAAAAGAGATATATATACATCTAGA 3605  
319 pPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysTyrPheL 336  
|:::|||||  
3606 TTTGGTACTTCACCATGAAGAGCTGGAGAGTATATGCTATAAATATTTTA 3655  
336 yLysGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352  
|:::|||||  
3656 AGGAACAGGAAGAAAATTTGTTTTATCTATTATTAAGTATCTTAATGAA 3705  
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369  
|:::|||||  
3706 TTTTATAGACTATAGAAATAAAGAATATGATGAACACCATCATATAG 3755  
369 rCysGlnLeuLeuPheLysLysAspGluSerThrAspGluIleGlyL 386  
|||||  
3756 TTGTCAGTTGCTTTTAAAAAAGATGAGAAAGTACTGATGATATAGGAT 3805  
386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402  
|||||  
3806 TGATGGTATTTCCTTTCTACGAATCTGGAGTTTACGTAAAAAGATAT 3855  
403 LysAspTyrPheCysIleSerLysTrpTyrLeuLysGluValLysArgLy 419  
|||||  
3856 AAAGATTAATTTTGTATATAAGTAAATGGTACTTAAAGAGGTAAGAA 3905  
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProLysAspG 436  
|:::|||||  
3906 ACCATATAAGTCAAAATTTGGGATGTAATTTGGCAGTTTATTCTCTAA 3955  
436 luGlyTrpThrGlu 440  
|||||  
3956 AAGGGTGGAGCTGAA 3969

seq\_name: gb\_ba:CBO242628

seq\_documentation\_block:

LOCUS CBO242628 1326 bp DNA linear BCT 03-JUN-2000  
DEFINITION Clostridium botulinum boNT/B gene, partial.

ACCESSION AJ242628

VERSION AJ242628.1 GI:4914467

KEYWORDS boNT/B gene; botulinum neurotoxin type B.

SOURCE Clostridium botulinum.

ORGANISM

Clostridium botulinum;  
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
Clostridium.

REFERENCE 1 (bases 1 to 1326)

AUTHORS Lalli,G., Herreros,J., Osborne,S.L., Montecucco,C., Rossetto,O. and

Schiaivo,G.

TITLE Functional characterisation of tetanus and botulinum neurotoxins

Binding domains

JOURNAL J. Cell. Sci. 112 (Pt 16), 2715-2724 (1999)

MEDLINE 99343691

REFERENCE 2 (bases 1 to 1326)

AUTHORS Schiaivo,G.

Direct Submission

TITLE Submitted (26-MAY-1999) Schiavo G., Molecular Neuropathobiology,

Imperial Cancer Research Fund, 44 Lincoln s Inn Fields, WC2A 3PX,  
UNITED KINGDOM

## FEATURES

source  
Location/Qualifiers  
1..1326  
/organism="Clostridium botulinum"  
/db\_xref="taxon:1491"  
/note="natural mutant"  
1..1326  
/gene="boNT/B"  
CDS  
<1..>1326  
/gene="boNT/B"  
/codon\_start=3  
/transl\_table=11  
/evidence=experimental  
/product="botulinum neurotoxin type B"  
/protein\_id="CAB43706.1"  
/db\_xref="GI:4914468"  
/db\_xref="SPTREMBL:Q9X708"  
translation="YNNNTILTEIFNFKNSKILNNILNLRVDRNNLIDLSGVGANVE  
VVDGVELNDKNQFKLTSSNSTSEIRVTQNONIIFNSMFLDFSVSEWIRIPKYKNDGION  
YIHNEYTIINCINKNSGWKISIRGNRIIWLTDINGKTSVFEYSIREDDISDYINRW  
FVTYNSDNKIEIKYQISYELKDFWGNPLMYKEYYFNAGKNNSYIKLKDDSSV  
IFNTELSQNAKEIYKISYELKDFWGNPLMYKEYYFNAGKNNSYIKLKDDSSV  
GELIRSKYNQNSNTINYRNLYIGEKFIIRKNSQSINDDIVRKEDYIYLDFFNSNR  
ENRVAYKDFKEEKLVLANIYDSNEFYKTIQIKREYDEQPTYSQCLLPKDEESTDE  
IGLIGTHREYESSIVLKDYNKVCISKWLVKEVRKRPYNPLNGCNWQFI"

BASE COUNT 573 a 105 c 201 g 447 t  
ORIGIN

## alignment\_scores:

Quality: 2097.00 Length: 430  
Ratio: 5.017 Gaps: 0  
Percent Similarity: 97.209 Percent Identity: 91.395

## alignment\_block:

US-09-910-186a-8 x CBO242628 ..

Align seg 1/1 to: CBO242628 from: 1 to: 1326

3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTy 19  
|||||  
36 AACAAATATATAGCAAAATTTAAATAATATATCTTAAATTTAAAGATA 85  
19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36  
|||||  
86 TAGGGATAAATAATTAATAGATTTATCAGGATATGGGCAATGTAGAGG 135  
36 alTyrAspGlyValIleLeuAsnAspLysAsnGlnPheLysLeuThrSer 52  
|||||  
136 TATATGATGGGTCGAGCTTAATCAATAAATCAATTTAAATTAACCTAGT 185  
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69  
|||||  
186 TCAACAACACAGCTAGATAGAGTACTGACTCAAAATCAGAATATCATATTAA 235  
69 nSerValPheLeuAspPheSerValSerPheThrIleArgIleProLysT 86  
|||||  
236 TAGTATGTTCTTGATTTTGTAGTTAGCTTTTGGATAGAAATACCTAAAT 285  
86 yRLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102  
|||||  
286 ATAAGATGATGGGTATACAAATTTATATTCATATGAATATACAAATTAAT 335  
103 AsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAsnAr 119  
|||||  
336 AATTCGATTAATAAATAATCTCGCTGGGAAATATCTATTAGGGGTAATAG 385  
119 gIleIleThrPheIleAspIleAsnGlyLysThrLysSerValPheP 136  
|||||  
386 GATAATATGGACTTAACTATATAAATGGAAACCAAAATCAGATATTT 435  
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrpPhe 152

|||||  
436 TTGAATATAGCATAGAGACATATATCAGACTATATAAATAGATGGTTT 485  
153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnG 169  
|||||  
486 TTTGTAACATATTAATAAATTCGATAACGCTAAATTTATATTAATGG 535  
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAla 186  
|||||  
536 TAAGCTAGAAATCAAAATATAGATATTTAAAGATATAGGAGAAGTTAT 585  
186 snGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202  
|||||  
586 ATGCTGAAATTAATTTAAATTAGATGCTGATATAGATAGAACAAATTT 635  
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219  
|||||  
636 ATTTGGATGAAATATTTTCAGTATTTTAAACAGAAATTAAGTCAATCAA 685  
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAsp 236  
|||||  
686 TATTAAGAAATATATAAATTCATCATATAGCGAATATTTAAAGATTT 735  
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252  
|||||  
736 TTTGGGGAATCCCTTAATGTACAATAAAGAAATATATATCTTTAATCG 785  
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGly 269  
|||||  
786 GGAATAAATAATTCATATATTAATAAAGAAAGATTCATCTGAGGTGA 835  
269 uIleIleuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyr 286  
|||||  
836 AATTTTAACAGCTAGCAAAATATATCAAAATTCCAATTTATATAAATA 885  
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSerAsnSer 302  
|||||  
886 GAAATTTATATATCGAGAAAAATTTATATAGAGAAAGTCAAAATCT 935  
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319  
|||||  
936 CAATCTATAATATGATATATAGTTAGAAAAGAAAGATATATATATCTAG 985  
319 pPhePheAsnLeuAsnGlnGluTrpArgValThrTyrLysTyrPheL 336  
|||||  
986 TTTTATTAATTCAAATCGAGAGTGGAGATATATGCTTATAAAGATTTA 1035  
336 ySLysGluGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352  
|||||  
1036 AGGAAGAGGAAAAAATTTGCTTTTAGCTAATATATATATGATCTTAAT 1085  
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369  
|||||  
1086 TTTTACAAAATATACAAAATAAAGAAATATATGATGAACCAACCAATAG 1135  
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGly 386  
|||||  
1136 TTGCAATTTACTTTTAAAAAAGATGAAGAAGTACTGATGAGATAGGAT 1185  
386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402  
|||||  
1186 TGATTTGTTATTCATCGTTTTCACGAATCTGCAATTTGTTATAAAGATT 1235  
403 LysAspTyrPheCysIleSerLysTrpTyrLeuLysGluValLysArg 419  
|||||  
1236 AAAAATTTATTTGTATAAAGTAAATGGTACTTAAAGAGAGTAAAGGAA 1285  
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIle 432  
|||||  
1286 ACCATATAAACCCCAATTTAGGGTGTAAATGGCAGTTTCAAT 1325  
seq\_name: gb\_pat:A69701

seq\_documentation\_block:  
LOCUS A69701 3509 bp DNA linear PAT 07-MAY-1999  
DEFINITION Sequence 19 from Patent WO9807864.  
ACCESSION A69701  
VERSION A69701.1 GI:4774314  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
unclassified.  
REFERENCE 1 (bases 1 to 3509)  
AUTHORS Foster,K.A., Quinn,C.P. and Shone,C.C.  
TITLE RECOMBINANT TOXIN FRAGMENTS  
JOURNAL Patent: WO 9807864-A 19 26-FEB-1998;  
FOSTER KEITH ALAN (GB)  
FEATURES  
Location/Qualifiers  
source 1..3509  
/organism="unidentified"  
/db\_xref="taxon:32644"  
1..>3509  
/product="unnamed"  
1..>3509  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAB42524.1"  
/db\_xref="GI:4774315"  
/translation="MPVTINNFNDPIDNNIIMPEPPFARGTGRIYKAFKITDRIW  
IIPERYFGYKPEDFNKSSGIFNRDCEYDYPDLNTNDKKNIFLOTMIKLFNRIKS  
PLGKLEMIITNGIPLGDRVPLEEFTNLTASVTAKLISNPGEVRKKGIFANLII  
FGPGVLNENETIDIGLONHEFASREGGGIMQKFCPEYVSFVFNVOENKGCASFNR  
GYFSDPALIMHELHVLHGLIGIKVDLPIVPNEKKFFMOSTDAIOAEELYTFFGGD  
PSIITPDSKSIYDKVLQNFRIQVDRNLKLVLCISDPNININIKKFKDKKKEVEDS  
EGKYSIDVESPDKLYKSLMFGFTETNIAENYKIKTRASYSDSLPPVKIKNLLDNEYI  
TIERGFNSDKDMEKRYGONKAIKQAYBEISKHELVAVYKIQMCKSVKAPGICIDVD  
NEDLEFTADKNSFDDLSKNERIENYQSNYIENDPPINELILDTDLISKLELSENT  
ESLDFNDVDPVVEKQPAKKIFDENTIFQIYLSQFPLDIRDISLSSDEDDALLFS  
NKVYSFSDMIDITANKVVEAGLFAGWQKIVNDFVIEANKSNMTMDKIADISLIVPYI  
GLAINVGNETAKGFENAFETAGASILEFPELLIPVVGAFLESYIDNKNKIYKI  
DUALTKREKMSDMYGLTIAQWLSTVNTQFYTIKEGMYKALNYOQALEEIKYRNI  
YSEKESKINIDFNIDNSKLEGINQADINNFINGCSVSLMKMKMPLAVERKLLDF  
DNTLKKNLNIDENKLYLGSAEYKSKVNYLKTIMPFDLSIYNDTILILENFYK  
NSEILNIIILNLRDKNLDLSIGYAKVEYDGVDELNMQFKLTSSANSKIRYTON  
QNIILNSVDFSVFWRIPRYKNDGIONIHNEYITLNCMKNNSGKWSIRGNRII  
WTLIDINGKTSVFVEDNREDISEYINRWFVTITNNLNNAKIYINGKLESNTDIKD  
IREVIANGELIKFKLDGIDTQFIMKYFSIFNTELSQSNIEERYKIQSYSEYKDEW  
GNPLMKEYVMFNAGNKNYSIKLKDKSPVGEILTRSKYNQNSYINVRDLYIGKEFI  
IRKSNQSINDIVRKEDIYI"  
BASE COUNT 1468 a 340 c 548 g 1153 t  
ORIGIN

alignment\_scores:  
Quality: 1667.00 Length: 318  
Ratio: 5.259 Gaps: 0  
Percent Similarity: 99.686 Percent Identity: 99.686

alignment\_block:

us-09-910-186a-8 x A69701

Align seg 1/1 to: A69701 from: 1 to: 3509

1 MetAlaAsnLysTyrAsnSerGluIleLeuAsnAsnIleleLeuAsnLe 17  
|||||  
2554 ATGTTTAAATAAATAATAGCGAAATTTAAATAATATTATCTTAAATTT 2603  
17 uArgTyrLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysV 34  
|||||  
2604 AAGATATAAGGATAATAATTTAATAGATTATCAGGATATGGGCAAGG 2653  
34 alGluValTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeu 50  
|||||  
2654 TAGAGGTATATGATGGAGTCGACGCTTAATGATAAAAAATCAATTAATTA 2703  
51 ThrSerSerAlaAsnSerLysLysIleArgValThrGlnAsnGlnAsnIleI 67

|||||  
2704 ACTAGTTCAGCAAAATAGTAGATTAGAGTGACTCAAAATCAGAAATATCAT 2753  
67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84  
|||||  
2754 ATTATAAGTGTGCTCCTGATTTTACGGTTAGCTTTTGGATAGAATAC 2803  
84 rOLysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThr 100  
|||||  
2804 CTAATAATAGGAATGATGGTATACAAAATATATATTCATATGAATATACA 2853  
101 IleIleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGI 117  
|||||  
2854 ATAATTAATGTATGAAAAATAATTCGGGCTGGAATAATATCTATTAGGG 2903  
117 yAsnArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerV 134  
|||||  
2904 TAATGAGTAATATGGACTTTAATTCATATAAATGGAACCAATCGG 2953  
134 aPhePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArg 150  
|||||  
2954 TATTTTTCGAATATAACATAAGAGAAGATATATCAGATATATAAATAGA 3003  
151 TrpPhePheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrI 167  
|||||  
3004 TGGTTTTTGTAACTATTACTAATAATTTCAATACGCTAAAAATTTATAT 3053  
167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184  
|||||  
3054 TAATGTAGCTAGATAAATCAACATATTAAGATATTAAGAGAGAGTTA 3103  
184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200  
|||||  
3104 TTGCTAATGGTGAATAAATATTTAAATTAGCTGCTATAGATAGAACA 3153  
201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGI 217  
|||||  
3154 CAATTATTGGATGAAATATTTTCAGTATTTTAAATACGGAATTAAGTCA 3203  
217 nSerAsnIleGluGluArgTyrIlysIleGlnSerTyrSerGluTyrLeuL 234  
|||||  
3204 ATCAAAATTTGAAGAAGAAGATATAAAATCAATCATATAGCGAATATTAA 3253  
234 yAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250  
|||||  
3254 AAGATTTTGGGGAATCCTTTAATGTCACATAAAGAATATATATATGTTT 3303  
251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProVa 267  
|||||  
3304 AATCGCGGGAATAAAATTCATATATTAACCTAAAGAAAGATTCACTGT 3353  
267 lGlyGluIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleA 284  
|||||  
3354 AGGTGAATTTTAAACAGTAGCAAAATATATCAAAATTTCTAAATATATAA 3403  
284 snTyrArgAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSer 300  
|||||  
3404 ATTATAGAGATTATATATATGGAGAAAATTTATATATAGAGAAAGTCA 3453  
301 AsnSerGlnSerIleAsnAspIleValArgLysGluAspTyrIleTf 317  
|||||  
3454 AATTCTCAATCTATAAATGATGATAGTTAGAAAAAGAGATTATATATA 3503  
317 rLeu 318  
|||||  
3504 TCTA 3507

seq\_name: gb\_pat:BD009886

seq\_documentation\_block:

LOCUS BD009886

DEFINITION Recombinant toxin fragments.

ACCESSION BD009886

VERSION BD009886.1 GI:18638259

KEYWORDS JP 2001502890-A/10.

SOURCE unidentified.

ORGANISM unidentified.

REFERENCE 1 (bases 1 to 3509)

AUTHORS Shone,C.C., Quinn,C.P. and Foster,K.A.

TITLE Recombinant toxin fragments

JOURNAL Patent: JP 2001502890-A 10 06-MAR-2001;

COMMENT MICROBIOLOGICAL RESEARCH AUTHORITY CAMR, THE SPEYWOOD LABORATORY LTD

OS unidentified

PN JP 2001502890-A/10

PD 06-MAR-2001

PF 22-AUG-1997 JP 1998510524

PR 23-AUG-1996 GB 9617671.4 13-DEC-1996 GB 9625996.5 PI

CLIFFORD CHARLES SHONE,CONRAD PADRAIG QUINN,KEITH ALAN FOSTER PC

C12N15/31,C12N1/21,C12P21/02,C07K14/33,A61K38/16,A61K39/08 CC

Strandedness: Double;

CC Topology: Linear;

FH Key Location/Qualifiers

FT CDS 1..3509.

source Location/Qualifiers

1..3509

/organism="unidentified"

/db\_xref="taxon:32644"

BASE COUNT 1468 a 340 c 548 g 1153 t

ORIGIN

alignment\_scores:

Quality: 1667.00 Length: 318

Ratio: 5.259 Gaps: 0

Percent Similarity: 99.686 Percent Identity: 99.686

alignment\_block:

US-09-910-186A-8 x BD009886 ..

Align seg 1/1 to: BD009886 from: 1 to: 3509

1 MetAlaAsnLysTyrAsnSerGluLeuAsnAsnIleLeuAsnLe 17

|||||

2554 ATGTTTAATAATATATACGGAATTTTAAATAATATTATCTTAAATTT 2603

17 uArgTyrLysAspAsnLeuIleAspLeuSerGlyTyrGlyAlaLysV 34

|||||

2604 AAGATATAAGGATAATAATTTATAGATTTATCAGGATATGGGCAAGG 2653

34 alGluValTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeu 50

|||||

2654 TAGAGGTATATGAGTGCAGCTTAATGATAAAATCAATTTAAATTA 2703

51 ThrSerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleI 67

|||||

2704 ACAGTTTCAGCAATAGTAGATAGAGTACTCAAAATCAGAAATATCAT 2753

67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84

|||||

2754 ATTTAATAGTGTCTTCATTTTACCGTTAGCTTTGGATAGAAATAC 2803

84 roLysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThr 100

|||||

2804 CTAATATAAGATGATGGTGATACAAAATTTATATTCATAATGAATACA 2853

101 IleIleAsnGlyMetLysAsnAsnSerGlyTrpLysIleSerIleArgGI 117

|||||

2854 ATAAATTAATGTTGATAAAATAATTCGGCTGGGAAATATCTATTAGGG 2903

117 YAsnArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerV 134

|||||

2904 TAATAGTAATATGACTTTTATTTATGATTAATGGAACCAATCCG 2953

134 alPheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArg 150

|||||

2954 TATTTTTGAATAACATAAGAGAAGATATATCAGAGTATATAAATAGA 3003

151 TrpPhePheValThrIleThrAsnAsnLeuAsnAlaLysIleTyrIl 167

|||||

3004 TGGTTTTTGTAACTATTACTAATAATTGAATACGCTAAATTTATAT 3053

167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184

|||||

3054 TAATGGTAACTGAATCAATACAGATATTAAAGATATAAGAGAAGTTA 3103

184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200

|||||

3104 TTGCTAATGGTGAATAATAATTAAATAGATGGTGATATAGAGAACA 3153

201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGI 217

|||||

3154 CAATTTATTTGGATGAATAATTTTCAATCAATCAATCAATCAATCAAT 3203

217 nSerAsnIleGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234

|||||

3204 ATCAATATTTGAAGAAGATATAAAATTCATCATATAGCGAATATTAA 3253

234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250

|||||

3254 AAGATTTTGGGAAATCCTTTAATGTACAATAAAGAAATATATATGTTT 3303

251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProVa 267

|||||

3304 AATCGGGGAATAAAATTCATATATTAACTAAAGAAAGATTCACTGT 3353

267 lGlyGluIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleA 284

|||||

3354 AGGTGAATTTTACACGACGACGACGACGACGACGACGACGACGACG 3403

284 snTyrArgAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSer 300

|||||

3404 ATTATAGAGATTTATATATTGGAGAAAATTTATTATAAGAAAGAAAGTCA 3453

301 AsnSerGlnSerIleAsnAspIleValArgLysGluAspTyrIleTy 317

|||||

3454 AATTCTCAATCTAATATGATATAGTTAGAAAAGAAAGATTTATATATA 3503

317 rLeu 318

|||||

3504 TCTA 3507

seq\_name: gb\_ba:CBONTG

seq\_documentation\_block:

LOGUS CBONTG 3937 bp DNA linear BCT 18-JAN-1994

DEFINITION C.botulinum BONT/G gene.

ACCESSION X74162

VERSION X74162.1 GI:441275

KEYWORDS bONT/G gene; botulinum neurotoxin.

SOURCE Clostridium botulinum.

ORGANISM Clostridium botulinum

Bacteria: Firmicutes; Bacillus/Clostridium group: Clostridiaceae;

Clostridium.

REFERENCE 1 (bases 1 to 3937)

AUTHORS Campbell,K.D.

TITLE Direct Submission

JOURNAL Submitted (13-JUL-1993) K.D. Campbell, AFRC Institute of Food Research, Reading Laboratory Dept of Microbiology, Earley Gate Whiteknights Road, Reading RG6 2EF, UK

REMARK revised by author 06-SEP-93

REFERENCE 2 (bases 1 to 3937)

AUTHORS Campbell,K., Collins,M.D. and East,A.K.

TITLE Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium argentinense) type G neurotoxin: genealogical comparison with other clostridial neurotoxins

JOURNAL Biochim. Biophys. Acta 1216 (3), 487-491 (1993)

MEDLINE 94092745

FEATURES Location/Qualifiers









203 IleTprMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219  
 6774 ATTGATGAATATTTTCAAGTGTCTTTTAAATACGGAATTAAGTCAATCAA 6823  
 219 nileGluGluArgTyrLysIleGlnSerTyrSerGlu 231  
 6824 TATTGAAGAAAGATATAAAATTCATCATATATAGCGAA 6860

seq\_name: gb\_pat:AR0000029

seq\_documentation\_block:

LOCUS AR0000029 1330 bp DNA linear PAT 04-DEC-1998

DEFINITION Sequence 22 from patent US 5736139.

ACCESSION AR0000029

VERSION AR0000029.1 GI:3962560

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 1330)

AUTHORS Kink,J.A., Thalley,B.S., Stafford,D.C., Firca,J.R. and Padhye,N.V.

TITLE Treatment of Clostridium difficile induced disease

JOURNAL Patent: US 5736139-A 22 07-APR-1998;

FEATURES

source

1. .1330

BASE COUNT 400 a 339 c 246 g 345 t

ORIGIN

alignment\_scores:

Quality: 831.50 Length: 450

Ratio: 2.607 Gaps: 10

Percent Similarity: 70.889 Percent Identity: 40.000

alignment\_block:

US-09-910-186a-8 x AR0000029 ..

Align seg 1/1 to: AR0000029 from: 1 to: 1330

4 LysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTyrLy 20  
 28 GAATACATCAAGAACATCATCAATACCTCCATCCATCGAACCCTGCGTAGCA 77  
 20 sAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluValT 37  
 78 ATCCATCACTGATCGACCTGCTCGCTGCTTCCAAATCAACATCG 127  
 37 yrAspGlyValGluLeuAsn.....AspLysAsnGlnPheLysLeuThr 51  
 128 GTTCTAAAGTTAACTTCGATCCGATCGACAGAAATCAGATCCAGCTGTC 177  
 52 SerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePh 68  
 178 AATCTGGAATCTTCCAAATCGAAGTTATCTCTGGAAGATGTCGTATA 227  
 68 eAsnSerValPheLeuAspPheSerValSerPheTprIleArgIleProL 85  
 228 CAACTTAATGACGAAACTTCTCCACCTCTCTGATCCGTATCCCGA 277  
 85 ystTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIle 101  
 278 AATACTTCAAC.....TCCATCTCTCTGAACAATGAATACACCATC 318  
 102 IleAsnCysMetLysAsnAsnSerGlyTprLysIleSerIleArgGlyAs 118  
 319 ATCAACTGCATGGAAGAAATCTGGTGGAAAGTATCTCTGAACATCGG 368  
 118 nArgIleIleTprThrLeuIleAspIleAsnGlyLysThrLysSerValP 135  
 369 TGAATCATCTGGACTGCGAGGACACTCAGGAATCAACACGCGGTGTG 418

135 hePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTpr 151  
 419 TATTCAATACTCTCAGATGATCAACATCTCTGACTACATCAATCGCTGG 468  
 152 PhePheValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrIl 167  
 469 ATCTTCGTTACCATCACCACAAATCGTCTGAATAACTCCAAATCTACAT 518  
 167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184  
 519 CAACGGCGCTGTATCGACCAACACCGCATCTCCAATCTGGGTAAACATCC 568  
 184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200  
 569 ACGCTTCTTAATAACATCATGTTCAAACTGGAGGTGTCTGACACTCAC 618  
 201 GlnPheIleTprMetLysTyrPheSerIlePheAsnThrGluLeuSerGl 217  
 619 CGCTACATCTGGATCAATACTTCAATCTGTTCGACAAAGAACTGAACGA 668  
 217 nSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234  
 669 AAAAGAAATCAAGAACCTGTACGACACCAAGTCCAAATCTGGTATCCCTGA 718  
 234 ysAspPheTprGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250  
 719 AAGACTTCTGGGTGACTACCTGCAGTAGCAGACAAACCGTACTACATGCTG 768  
 251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLys..... 263  
 769 AATCTGTACGATCCGACAAATACGTTGACGTCACAAATAGTAGGTATCCG 818  
 264 .....AspSerProValGlyGluIleLeuThrArgSerL 275  
 819 CGTTACATGATCTGAAGGTCCGCTGCTGTATGACTACCAACA 868  
 275 ystTyrAsnGlnAsnSerLysTyrIleAsnTyrArgAspLeuTyrIleGly 291  
 869 TCTACTGAACCTCTCC.....CTGTACCGTGT 897  
 292 GluLysPheIleIleArgArgLysSerAsnSerGlnSerIleAsnAspAs 308  
 898 ACCAAATTCATCATCAAGAAATACGCGTCTGTAAAC.....AAGGACAA 941  
 308 pIleValArgLysGluAspTyrIleTyrLeuAspPheAsnLeuAsnG 325  
 942 TATCGTTCGCAACAATGATCGTGATATACATCAATGTTGATTAGTAACA 991  
 325 InGluTprArgValTyrThrTyrLysTyrPheLysLysGluGluGluLys 341  
 992 AAGAATACCGTCTGGCTACCAATGCTTCTCAGGCTGGTGTAGAAAAGATC 1041  
 342 LeuPheLeuAlaProIleSerAspSerAspGluLeuTyrAsnThrIleGl 358  
 1042 TTGCTGCTCTGGAATCCCGAGCTGTGTAATCTCTCAGGTAGTTGT 1091  
 358 nIleLysGluTyrAspGluGlnPro...ThrTyrSerCysGlnLeuLeuP 374  
 1092 AATGAATCCAAAGAACGACGAGGTATCACTAACAAATGCCAAATG.... 1137  
 374 heLysLysAspGluSerThrAspGluIleGlyLeuIleGlyIleHis 390  
 1138 ..AATCTCGAGGACAAATGTTACGATATCGGTTTCATCGTTTCCAC 1185  
 391 ArgPheTyrGluSerGlyIleValPheGluGluTyrLysAspTyrPheCy 407  
 1186 CAGTTCAACAATATCGCT.....AAACTGGT 1211  
 407 sIleSerLysTprTyrLeuLysGluValLysArgLysProTyrAsnLeuL 424  
 1212 TGCCTCCAACTGGTACATCGTCAGATCGAAGCTTCC.....TCTCGCA 1255  
 424 ysLeuGlyCysAsnTrpGlnPheIleProLysAspGluGlyTprThrGlu 440

1256 CTCTGGGTTGCTCTTGGGAGTTTCATCCCGGTTGATGACGGTTGGGGTGAA 1305



OM of: US-09-910-186A-8 to: N\_Geneseq\_032802:\* out\_format : pfs  
Date: Sep 2, 2002 4:04 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=Cgn2\_1/USPTO\_spool/US09910186/runat\_29082002.134726.16035/app\_query.fasta\_1.505  
-DB=N\_Geneseq\_032802 -QFMT=fastcap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62  
-TRANS=human40.cdi -LISP=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09910186.@CGL\_1\_178 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT\_THREADS=1

Search information block:  
Query: US-09-910-186A-8  
Query length: 440  
Database: N\_Geneseq\_032802:\*  
Database sequences: 1736436  
Database length: 858457221  
Search time (sec): 195.180000

score\_list:

Sequence	Strd Orig	zScore	eScore	Len	! Documentation
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87213	...	2350.00	3639.26	2.3e-19	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87213	...	2350.00	3639.26	2.3e-19	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30581	...	2340.00	3632.40	2.0e-19	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30579	...	2333.00	3603.09	2.4e-19	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87212	...	2227.00	3448.06	1.0e-19	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30580	...	2102.00	3252.64	8.0e-17	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV26289	...	1667.00	2568.94	9.6e-13	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87217	...	1168.50	1802.99	4.4e-92	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87217	...	1168.50	1802.99	4.4e-92	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30596	...	1168.50	1801.79	5.2e-92	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87220	...	831.50	1279.45	6.4e-63	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216	...	831.50	1279.40	6.5e-63	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216	...	831.50	1279.38	6.5e-63	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAZ87216	...	831.50	1279.36	6.5e-63	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216	...	831.50	1279.36	6.5e-63	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87216	...	831.50	1279.34	6.5e-63	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87216	...	831.50	1279.30	6.6e-63	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30576	...	831.50	1279.21	6.6e-63	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30572	...	831.50	1278.87	6.9e-63	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30572	...	831.50	1278.87	6.9e-63	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30575	...	831.50	1277.97	7.8e-63	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87218	...	831.50	1273.44	1.4e-62	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT:AAZ87244	...	831.50	1269.50	2.3e-62	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87216	...	831.50	1267.50	3.0e-62	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216	...	762.50	1172.11	6.1e-54	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:AAZ87216	...	732.00	1124.98	2.6e-54	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216	...	732.00	1124.84	2.6e-54	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216	...	732.00	1124.83	2.6e-54	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216	...	732.00	1124.81	2.6e-54	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87216	...	728.00	1118.59	5.9e-54	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30593	...	723.00	1109.87	1.8e-53	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30585	...	693.50	1033.95	6.5e-51	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216	...	692.00	1052.40	2.9e-51	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216	...	688.50	1057.48	1.5e-50	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216	...	688.50	1056.64	1.7e-50	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216	...	688.50	1056.64	1.7e-50	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30584	...	676.00	1036.81	2.1e-49	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30591	...	615.50	942.75	3.7e-44	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216	...	612.50	938.70	6.1e-44	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT:AAQ12121	...	594.00	910.05	2.4e-43	

seq_name:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87213
seq_documentation_block:	
ID	AAZ87213 standard; DNA; 1341 BP.
XX	
AC	AAZ87213;
XX	
DT	08-MAY-2000 (first entry)
XX	
DE	DNA encoding synthetic BoNT serotype B (BoNTB) Hc fragment.
XX	
KW	Botulinum neurotoxin; heavy chain; BoNT; serotype B;
KW	C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KW	VEE; botulism; vaccine; diagnosis; drug screening; ds.
XX	
OS	Clostridium botulinum.
OS	Synthetic.
XX	
FX	Key
FX	Location/Qualifiers
FT	10..1333
FT	/tag= a
FT	/product= "Synthetic botulinum neurotoxin serotype B
FT	(BoNTB) heavy chain C-terminal fragment (Hc)"
XX	
PN	WO200002524-A2.
XX	
PD	20-JAN-2000.
XX	
PF	09-JUL-1999; 99WO-US15570.
XX	
PR	10-JUL-1998; 98US-0092416.
PR	12-MAY-1999; 99US-0133870.
XX	
PA	(USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
PI	Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
DR	WPI: 2000-160827/14.
DR	P-PSDB; AAY77135.
XX	
PT	Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum toxin serotypes A-G, is used for inducing an immune response against botulinum -
PT	
PS	Disclosure; Page 39-40; 54pp; English.
XX	
CC	The invention relates to novel vaccines that induce a protective immune response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant DNA construct comprising a vector, and at least one nucleic acid fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-G. In preferred embodiments of the invention, the vector is a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of this vector results in the production of large amounts of a protein encoded by a sequence cloned into the replicon. The constructs are used to produce vaccines against botulism. The proteins can also be used as diagnostic tools for the diagnosis of botulism. The transformed host cells can be used to analyse the effectiveness of drugs and agents which inhibit toxin effects. The vaccine currently used against botulism is dangerous and expensive to produce, and contains formalin, which is very painful for the recipient. Also, the vaccine is incomplete, in that only 5 of the 7 serotypes are represented in the formulation. The novel vaccine overcomes these problems, as it is easily purified, and available in large quantities. It is also expressed in the lymph nodes for a better immune response. Sequences AAZ87212-287217 represent synthetic DNA sequences encoding BoNT Hc fragments used in the present invention. These were optimised for codon usage for expression in yeast.

xx

SQ Sequence 1341 BP; 442 A; 332 C; 235 G; 332 T; 0 other;

## alignment\_scores:

Quality: 2350.00 Length: 440  
Ratio: 5.341 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-910-186a-8 x AA287213

Align seg 1/1 to: AA287213 from: 1 to: 1341

1 MetAlaAsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLe 17  
10 ATGGCCACAATAACAAATCCGAAATCTCGAACAATATCATCTCGAACC 59  
17 uArgTyrLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysv 34  
60 GGGTTACAAAGACAACAATCTGATCGATCTGTCTGGTTACGGTGCTAAAG 109  
34 aGluValTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeu 50  
110 TTGAAGTATACGACGGTGTGAACCTGAATGACAAAGACCACTTCAAAC 159  
51 ThrSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleI 67  
160 ACCTTCCGCTTAACCTTAAGATCCGCTTACTCAGAATCAGAATCAT 209  
67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84  
210 CTTCAACTCCGTATTCTCCGACTCTCTGTTCTCTTCTGGATTCGTATCC 259  
84 rLysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThr 100  
260 CGAAATACAAAGAACGCGTATCCAGAAATTATACCCACAATGAATACACC 309  
101 IleIleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgG 117  
310 ATCATCAACTGCAGAGAAATAACTCTGGTGGAAAGATCTCCATCCGCGG 359  
117 yAsnArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerV 134  
360 TAACCGTATCATCTGGACTCTGATCGATATCAACGGTAAGACCAATCTG 409  
134 aPhePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArg 150  
410 TATTTCTCGAATACAAACATCCGTGAAGACATCTCTGAATACATCAATCGC 459  
151 TrpPhePheValThrIleThrAsnAsnLeuAsnAlaLysIleTyrI 167  
460 TGGTCTTCGTTTACCATTACCAATTAACCTGAACATGCTAAATCTCAT 509  
167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184  
510 CAACGGTAACTGGAATCTAATACCGACATCAAGACATCCGTGAAGTTA 559  
184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200  
560 TCGCTTAACGGTGAATCATCTCAACCTGGACGGTGACATCGATCGTACC 609  
201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerG 217  
610 CAGTTCATCTGGATGAATATCTTCCATCTTCAACCCGACCTGCTCA 659  
217 nSerAsnIleGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234  
660 GTCCAATATCGAAGAACGGTACAAGATCCAGTCTTACTCCGAATACCTGA 709  
234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250  
710 AAGACTTCTGGGGTAATCCGCTGATGTACAACAAAGAAATACTATATGTT 759

251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProVa 267  
760 AATGCTGTGTAAACAACACTTTACATCAAACTGAAGAAAGACTCTCCGCT 809  
267 lGlyGluIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleA 284  
810 TGGTGAATCTCTGACTCGTTCCAATACAAACAGAACTCTAAATACATCA 859  
284 snTyrArgAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSer 300  
860 ACTACCGGACCTGTACATCGGTGAAGATTCATCATCGTCGCAATCT 909  
301 AsnSerGlnSerIleAsnAspIleValArgLysGluAspTyrIleTyr 317  
910 AACTCTCAGTCCATCATGATGACATCGTACGTAAGAAGACTACATCTA 959  
317 rLeuAspPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysT 334  
960 CCTGGACTTCTTCAACCTGAATCAGGAATGGCGTGTATACACTACAAGT 1009  
334 yrPheLysLysGluGluLysLeuPheLeuAlaProIleSerAspSer 350  
1010 ACTTCAAGAAAGAGAGAAAGCTTTTCTCGTCCGATCTCTGATTC 1059  
351 AspGluLeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProth 367  
1060 GACGAACTTACACACCATCCAGATCAAGAATACGACGACACCGGAC 1109  
367 rTyrSerCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluI 384  
1110 CTACTCTTCCAGCTGCTGTTCAGAAAGATGAAGAATCTACTACGAAA 1159  
384 leGlyLeuIleGlyIleHisArgPheTyrGluSerGlyIleValPheGlu 400  
1160 TCGGTCTGATCGGTATCCACCGTTTCTACGAATCTGGTATCGTATCGAA 1209  
401 GluTyrLysAspTyrPheCysIleSerLysTyrTrpTyrLeuLysGluVal 417  
1210 GAATACAAGACTACTTCTGCATCTCCAAATGTTACTGACGAAAGTTAA 1259  
417 sArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProL 434  
1260 ACGCAAAACCGTACAACCTGAAACTGGTTGCAATGGCAGTTTCATCCCGA 1309  
434 ysAspGluGlyTrpThrGlu 440  
1310 AAGCGAAGGTGGACCGAA 1329  
seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAA54485  
seq\_documentation\_block:  
ID AAA54485 standard; DNA; 1341 BP.  
XX  
AC AAA54485;  
XX  
DT 11-APR-2001 (first entry)  
XX  
DE Botulinism toxin heavy chain C-terminal coding sequence (serotype B).  
XX  
KW Botulinism; toxin; neurotoxin; heavy chain; recombinant expression;  
KW recombinant vector; antigen; immune response; vaccine; bacterium;  
KW infection; ds.  
XX  
OS Synthetic.  
OS Clostridium botulinum.  
XX  
FH Key Location/Qualifiers  
FT CDS 10..1332  
FT /tag= a  
FT /product= H\_C peptide fragment  
XX  
PN WO200067700-A2.



1110 CTACTCTGCCAGCTGCTGTTCAAGAAAGATGAAGAATCTACTGACGAAA 1159  
 384 leGlyLeuileGlyIleHisArgPheTyrGluSerGlyIleValPheGlu 400  
 1160 TCGGTCTGATCGGTATCCACCGTTTCACGAATCTGGTATCGTATTCGAA 1209  
 401 GluTyrLysAspTyrPheCysIleSerLysTrpTyrLeuLysGluValLy 417  
 1210 GAATACAAAGACTACTCTGCATCTCCAATGGTACCTGAAGGAAGTTAA 1259  
 417 sArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProL 434  
 1260 ACCGAAACCGTACAACTGAACTGGGTGCAATGGCAGTTTCATCCGA 1309  
 434 ysAspGluGlyTrpThrGlu 440  
 1310 AAGACGAAGGTGGACCGAA 1329

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30581

seq\_documentation\_block:

ID AAV30581 standard; DNA; 1547 BP.

AAV30581;

07-DEC-1998 (first entry)

Clostridium botulinum toxin B fragment C gene in pETHisB.

Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen; botulism; BotB; ds.

Clostridium botulinum serotype B Danish strain.

Synthetic.

Key Location/Qualifiers

CDS 108..1526

/\*tag= a

W09808540-A1.

05-MAR-1998.

28-AUG-1997; 97WO-US15394.

28-AUG-1996; 96US-0704159.

(OPHI-) OPHIDIAN PHARM INC.

Thalley BS, Williams JA;

WPI; 1998-230234/20.

P-PSDB; AAW68394.

Host cell containing recombinant expression vector encoding Clostridium botulinum type B or E toxin - useful to treat humans and other animals at risk of intoxication with clostridial toxin

Example 35; Page 303-305; 428pp; English.

This is the DNA sequence of the Clostridium botulinum serotype B (Danish strain) toxin fragment C gene contained in plasmid pETHisB. The encoded fragment C polypeptide (see AAW68394) has a His-tagged N-terminal extension. The vector was used to express native (i.e. non-fusion) soluble C fragment in Escherichia coli host cells. The invention relates to recombinant proteins derived from C. botulinum toxins. Methods are provided which allow for the isolation of soluble recombinant proteins free of significant endotoxin contamination. Preferred hosts for production of recombinant proteins are E. coli, insect cells and yeast cells. The recombinant toxins are used as immunogens for the production of vaccines and antitoxins that are useful in the treatment of humans and animals at risk of intoxication with clostridial toxin.

XX  
 SQ Sequence 1547 BP; 634 A; 148 C; 263 G; 502 T; 0 other;  
 alignment\_scores:  
 Quality: 2340.00 Length: 440  
 Ratio: 5.330 Gaps: 0  
 Percent Similarity: 99.773 Percent Identity: 99.545  
 alignment\_block:  
 US-09-910-186a-8 x AAV30581 ..  
 Align seg 1/1 to: AAV30581 from: 1 to: 1547  
 1 MetAlaAsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLe 17  
 204 ATGTTTAAATAATATATAGCGAAATTTTAAATAATATATATCTTAAATTT 253  
 17 uArgTyrLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysV 34  
 254 AAGATATAAGGATAATAATTTAATAGATTTATCAGGATATGGGCAAGG 303  
 34 alGluValTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeu 50  
 304 TAGAGGTATATGATGGAGTCGAGCTTAATGATAAAATCAATTTAAATTA 353  
 51 ThrSerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleI 67  
 354 ACTAGTTCAGCAATATAGTAAGATTAGAGTGACTCAAAATCAGATATCAT 403  
 67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84  
 404 ATTTAATAGTGTCTTCCTTGATTTTAGCGTTAGCTTTTGGATAAGAATAC 453  
 84 roLysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThr 100  
 454 CTAATATAAGAATGATGGTATACAAAATTTATTCATAATGAATATACA 503  
 101 IleIleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgG1 117  
 504 ATAATTAATGTTGTAATAATAATTCGGCTGGAAAAATATCTATTAGGGG 553  
 117 yAsnArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysServ 134  
 554 TAATAGGATAATATGGACTTTAATTGATATAAATGGAACCAAAATCGG 603  
 134 alPhePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArg 150  
 604 TATTTTGTGAATATAACATAAGAGAAGATATATCAGAGTATATAAATAGA 653  
 151 TrpPhePheValThrIleThrAsnAsnLeuAsnAlaLysIleTyrI1 167  
 654 TGGTTTTTGTAACTATTACTAATAATTTGAATTAACGCTAAAAATTTATAT 703  
 167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184  
 704 TAATGGTAAGCTAGAAATCAAAATACAGATATTTAAAGATATAAGAGAAGTTA 753  
 184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200  
 754 TTGCTAATGGTGAATAATATTTAAATTTAGATGGTATATAGATAGAACAA 803  
 201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerG1 217  
 804 CAATTTATTTGGATGAATATTTTCAGTATTTTAAATACGGAATTAAGTCA 853  
 217 nSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234  
 854 ATCAAAATATTGAGAAAGATATAAAATTTCAATCATATATAGCGAATATTTAA 903  
 234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250  
 904 AAGATTTTGGGGAAATCCCTTTAATGTACAAATAAAGAAATATATTATGTTT 953



251 AsnAlaGlyAsnSerTyrIleLysLeuLysLysAspSerProVa 267  
|||||  
954 AATGCGGGAATAAAATTCATATATTAACCTAAAGAAAGATTCACCTGT 1003  
|||||  
267 IclYGlulLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleA 284  
|||||  
1004 AGGTCAAAATTTAAACACGTAGCAATATAATCAAAATCTTAATATATAA 1053  
|||||  
284 sNtyrArgAspLeuTyrIleGlyGluLysPheIleIleArgLysSer 300  
|||||  
1054 ATTATAGAGATTATATATTGGAGAAAATTTATTATAAGAAAGCA 1103  
|||||  
301 AsnSerGlnSerIleAsnAspIleValArgLysGluAspTyrIleTy 317  
|||||  
1104 RATTCTCAATCTATAAATGATGATATAGTAGAGAAAGAGATTATATA 1153  
|||||  
317 rLeuAspPheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysT 334  
|||||  
1154 TCTAGATTTTAAATTTAAATCAAGAGTGGAGATATATACCTATAAT 1203  
|||||  
334 YrPheLysLysGluGluLysLeuPheLeuAlaProIleSerAspSer 350  
|||||  
1204 ATTTTAAGAAAGAGGAGAAAATTTGTTTATAGCTCCCTATAAGTGATTCT 1253  
|||||  
351 AspGluLeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProth 367  
|||||  
1254 GATGAGTTTACAACTACTACAAATAAAGAAATATGATGAACGCCAAC 1303  
|||||  
367 rTyrSerCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluI 384  
|||||  
1304 ATATAGTTGTCTAGTTGCTTTTAAAAAAGATGAAGAAAGTACTGATGAGA 1353  
|||||  
384 leGlyLeuIleGlyIleHisArgPheTyrGluSerGlyIleValPheGlu 400  
|||||  
1354 TAGGATGATTGGTATTTCATCTCTTCTACGAATCTGGAATGTATTGAA 1403  
|||||  
401 GluTyrLysAspTyrPheCysIleSerLysTyrTyrLeuLysGluVally 417  
|||||  
1404 GAGTATAAGATTATTTTGTATAGTAAGTAAAGTACTTAAAGAGGTAA 1453  
|||||  
417 sArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProL 434  
|||||  
1454 AAGGAACCATATAATTTAAATTTGGGATGTAATTTGGCAGTTTATTCCTA 1503  
|||||  
434 ysAspGluGlyTrpThrGlu 440  
|||||  
1504 AAGATGAAGGGTGGACTGAA 1523  
|||||

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV30579

seq\_documentation\_block:  
ID - AAV30579 standard; DNA; 3876 BP.  
XX AC AAV30579;  
XX XX  
DT 07-DEC-1998 (first entry)  
XX XX  
DE Clostridium botulinum type B toxin gene from Danish strain.  
XX XX  
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;  
KW botulism; BotB; ds.  
XX XX  
OS Clostridium botulinum serotype B Danish strain.  
XX XX  
PN W09808540-A1.  
XX XX  
PD 05-MAR-1998.  
XX XX  
PF 28-AUG-1997; 97W0-US15394.  
XX XX  
PR 28-AUG-1996; 96US-0704159.  
XX XX

(OPHI-) OPHIDIAN PHARM INC.  
XX Thalley BS, Williams JA;  
XX WPI; 1998-230234/20.  
DR P-PSDB; AAW68392.  
XX Host cell containing recombinant expression vector encoding  
PT Clostridium botulinum type B or E toxin - useful to treat humans  
PT and other animals at risk of intoxication with clostridial toxin  
XX  
PS Example 35; Page 291-296; 428pp; English.  
XX  
CC This is the coding region of the Clostridium botulinum serotype B  
CC (Danish strain) toxin gene that codes for a 1291-amino acid  
CC polypeptide (see AAW68392). The C fragment (see AAW68394) of the  
CC B toxin has been expressed as histidine-tagged protein in Escherichia  
CC coli host cells. The invention relates to C. botulinum recombinant  
CC toxin polypeptides. Methods are provided which allow for the  
CC isolation of soluble recombinant proteins free of significant  
CC endotoxin contamination. Preferred hosts for production of the  
CC recombinant proteins are E. coli, insect cells and yeast cells.  
CC The recombinant proteins are used as immunogens for the production  
CC of vaccines and antitoxins that are useful in the treatment of  
CC humans and animals at risk of intoxication with clostridial toxin.  
XX  
SQ Sequence 3876 BP; 1612 A; 370 C; 617 G; 1277 T; 0 other;

alignment\_scores:  
Quality: 2333.00 Length: 440  
Ratio: 5.326 Gaps: 0  
Percent Similarity: 99.545 Percent Identity: 99.318  
alignment\_block:  
US-09-910-186A-8 x AAV30579 ..  
Align seg 1/1 to: AAV30579 from: 1 to: 3876

1 MetAlaAsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLe 17  
|||  
2554 ATGTTTAAATAATAATAAGCGAAATTTTAAATAATATATATCTTAAATTT 2603  
|||  
17 uArgTyrLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysV 34  
|||||  
2604 AAGATATAGGATATATATTAATGATGATTTATCAGGATATGGGCAAGG. 2653  
|||  
34 alGluValTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeu 50  
|||||  
2654 TAGAGTATATGATGAGTCGAGCTTAATGATATAAAATCAATTTAAATTA 2703  
|||  
51 ThrSerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleI 67  
|||||  
2704 ACTAGTTTCAGCAATAGTAAGATTAGAGTGACTCAAAATCAGAAATATCAT 2753  
|||  
67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84  
|||||  
2754 ATTTAATAGTGTCTTCCTGATTTTAGCCTTAGCCTTTTGGATAAAGATAC 2803  
|||  
84 rOLysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThr 100  
|||||  
2804 CTAATATAAAGATCATGCTATACAAAATATATTCATTAATGAATATACA 2853  
|||  
101 IleIleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgG 117  
|||||  
2854 ATAATTAATTTGATGAAAAATAATTCGGCTGAAAAATATCTATTAGGGG 2903  
|||  
117 yAsnArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerV 134  
|||||  
2904 TAATAGGATTAATAGGACTTTTAAATGATATAATGGAACCAATTCGG 2953  
|||  
134 alPhePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArg 150  
|||||

```
2954 TATTTTTTGAATATAACATAAGAGAGATATATACAGAGTATATAAATAGA 3003
151 TrpPheValThrIleThrAsnAsnLeuAsnAlaLysIleTyrIle 167
|||||
3004 TGGTTTTTGTAACTATTACTAATAATTGTAATACGCTAAATTTATAT 3053
167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184
|||||
3054 TAATGGTAAGCTAGAAATCAATACAGATATTAAGATATAAGAGAGTTA 3103
184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
|||||
3104 TTGCTAATGGTGAATAATATTTAAATAGATGTGATATAGATAGAACA 3153
201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerG 217
|||||
3154 CAATTAATTTGGATGGAATATTTTCAGTATTTTAAATACGGAATTAAGTCA 3203
217 nSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234
|||||
3204 ATCAAAATATTGAAGAAAGATATAAAATTCATCATATAGCGAATATTAA 3253
234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
|||||
3254 AAGATTTTTGGGAAATCCCTTATGTACAAATAAAGAAATATTATATGTT 3303
251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProva 267
|||||
3304 AATGCGGGGAATAAAATTCATATATTAACTAAAGAAAGATTCACCTGT 3353
267 lGlyGluIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleA 284
|||||
3354 AGGTGAATTTTAAACACGTAGCAATATATCAAAATTCATAATATATAA 3403
284 snTyrArgAspLeuTyrIleGlyLysPheIleIleArgArgLysSer 300
|||||
3404 ATTATAGAGATTTTATATTGGGAAAAATTTATTATAAGAAAGAAAGTCA 3453
301 AsnSerGlnSerIleAsnAspAspIleValArgLysGluAspTyrIleTy 317
|||||
3454 AATTCCTCAATCATAAATGATGATATAGTATAGTAAAGAAAGATATATA 3503
317 rLeuAspPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLys 334
|||||
3504 TCTAGATTTTTTTAAATTAATCAAGAGTGGAGAGTATATACCTATAAAT 3553
334 yrPheLysLysGluGluGluLysLeuPheLeuAlaProIleSerAspSer 350
|||||
3554 ATTTTAAAGAAAGAGAAAGAAATTTGTTTAGCTCCTATAAGTGATCT 3603
351 AspGluLeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProTh 367
|||||
3604 GATGAGTTTACATATATACATAATAAAGAAATATCATCAACAGCCAC 3653
367 rTyrSerCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluI 384
|||||
3654 ATATAGTTGTGCTGCTGCTTTTAAAAAAGATGAAGAAAGTACTGATCAGA 3703
384 leGlyLeuIleGlyIleHisArgPheTyrGluSerGlyIleValPheGlu 400
|||||
3704 TAGGATGATGGTGGTATTCATCGTTTCACGAATCTGGAATGTGATTGAA 3753
401 GluTyrLysAspTyrPheCysIleSerLysTyrTyrLysLysGluVally 417
|||||
3754 GAGTATAAGATATTATTTTGTATAAAGTAAATGGTACTTAAAGAGGTAA 3803
417 sArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProL 434
|||||
3804 AAGGAAACCATATAAATTAATAATGGGATGTAATGGCAGCTTTATTCCTA 3853
434 ysAspGluGlyTyrThrGlu 440
|||||
3854 AAGATGAAGGGTGGACTGAA 3873
```

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAA54589
seq_documentation_block:
ID AAA54589 standard; DNA; 1347 BP.
```

```
XX AAA54589;
AC
DT 11-APR-2001 (first entry)
XX
DE Sequence encoding botulinum toxin C fragment (serotype B).
XX
KW Botulinum toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection; ds.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US12890.
XX
PR 12-MAY-1999; 99US-0133865.
PR 12-MAY-1999; 99US-0133866.
PR 12-MAY-1999; 99US-0133867.
PR 12-MAY-1999; 99US-0133868.
PR 12-MAY-1999; 99US-0133869.
PR 12-MAY-1999; 99US-0133873.
PR 29-JUL-1999; 99US-0146192.
XX
(PUSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
WPI; 2001-016048/02.
XX
New nucleic acids encoding the carboxy- or amino-terminal portions of
the heavy chain of botulinum neurotoxin of serotype A-G, useful as
vaccine against botulinum
XX
Example 8; Page 38-39; 73pp; English.
XX
Botulinum neurotoxins are translated as a single 150 kDa polypeptide
chain and then posttranslationally nicked, forming a dichain
consisting of a 100 kDa heavy chain and a 50 kDa light chain which
remain linked by a disulfide bond. Nucleic acids encoding the
carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
chain of botulinum neurotoxin (BoNT) can be used in recombinant
expression vectors and expressed in transformed cells to produce
peptide antigens useful for eliciting an immune response to give
protective immunity against botulinum neurotoxin, which causes
botulism. The nucleic acids are expressible in a recombinant
organism such as Escherichia coli or Pichia pastoris. The use
of recombinant nucleic acids are advantageous since it eliminates
the need to culture large quantities of hazardous toxin-producing
bacterium. Production yield from the genetically engineered product
is also high and cost of production is lower. The nucleic acids can
be derived from Clostridium botulinum serotypes A-G.
XX
SQ Sequence 1347 BP; 439 A; 337 C; 238 G; 333 T; 0 other;
```

```
alignment_scores:
Quality: 2227.00      Length: 438
Ratio: 5.108          Gaps: 0
Percent Similarity: 99.543      Percent Identity: 99.315

alignment_block:
US-09-910-186A-8 x AAA54589 ..
```

Align seg 1/1 to: AAA54589 from: 1 to: 1347

```
3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleLeuAsnLeuArgTy 19
|||||
10 AACAAATACAAATCCGAAATCCGTAACAAATATCATCTCGAACCTGGTTA 59
|||||
19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36
|||||
60 CAAAGACAACAATCTGATCGATCTGCTGCTTACGGTGCTAAAGTTGAAG 109
|||||
36 alTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeuThrSer 52
|||||
110 TATACACCGGTGTGTAAGTCAATGCAAGAACCAGTTCAAATGACCTCT 159
|||||
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
|||||
160 TCCGCTAACTCTAAGATCCGTTACTCAGATCAGAACATCATCTTCAA 209
|||||
69 nSerValPheLeuAspPheSerValSerPheTrpIleArgIleProLysT 86
|||||
210 CTCGGTATCTCTGACTTCTCTGTTCTCTGATCCGATCCCGAAAT 259
|||||
86 yLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
|||||
260 ACAAGAACACGCGTATCCAGATTCATCCACAAATGAATACACCATCATC 309
|||||
103 AsnCysMetLysAsnAsnSerGlyTyrLysIleSerIleArgGlyAsnAr 119
|||||
310 AACTGATGAAGAATAACTCTGTTGGAGATCTCCATCCGCGGTAAACCG 359
|||||
119 gIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPheP 136
|||||
360 TATCATCTGGACTCTGATGATATCAACGGTAAGACCAATCTGTATTCT 409
|||||
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrpPhe 152
|||||
410 TCGAATACAAACATCCGTTGAAGACATCTCGAATACATCAATCGCTGTT 459
|||||
153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnG 169
|||||
460 TTCGTTACCATCAACATACCTTGAACATGCTTAAATCTTACATCAACGG 509
|||||
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAlaA 186
|||||
510 TAAACTGGAATCTAATACCGACATCAAGACATCCGCGAAGTTATCGCTA 559
|||||
186 snGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|||||
560 ACGGTGAAATCATCTTCAAACTGGACGCTGACATCGATCGTACCCAGTTC 609
|||||
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||
610 ATCTGGATGAATACCTCTCCATCTTCAACACCGAAGTGTCTCAGTCCAA 659
|||||
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAspP 236
|||||
660 TATCGAAGAACGCTACAGATCCAGTCTTACTCCGAATACCTGAAAGACT 709
|||||
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||
710 TCTGGGGTAATCCGCTGATGTACAACAAGAATACTATATGTTCAATGCT 759
|||||
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyG 269
|||||
760 GGTAACAAGAATCTTATACATCAAACTGAAAGAAAGACTCTCCGGTTGGTGA 809
|||||
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||
810 ATCCTGACTCGTCCAAATACAAACAGAACTCTTAATATACATCACTACC 858
|||||
286 rgAspLeuTyrIleGlyGluLysPheIleIleArgLysSerAsnSer 302
|||||
859 GCGACCTGTACATCGGTGAAAGTTTCATCATCGTCCGCAAAATCTAACTCT 908
|||||
```

```
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319
|||||
909 CAGTCCATCAATGA.GACATCGTAGCTAAGAAGACTACATCTACCTGGA 957
|||||
319 pPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysTyrPheL 336
|||||
958 CTCTCTCAACCTGAATCA.GAATGGCGTGTATACACCTACAAGTACTTCA 1006
|||||
336 yLysGluGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
|||||
1007 AGAAAGAAGAAGAAAGCTTTCTTGCTCCGATCTCTGATCCGACGAA 1056
|||||
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
|||||
1057 CTCTACAACACCATCCAGATCAAGAATAGCAGACAGACCGGACTACTC 1106
|||||
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
|||||
1107 TTGCCAGCTGCTGTTCAAGAAAGATGAAGAATCTACTGACGAAATCGGTC 1156
|||||
386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
|||||
1157 TGATCGGTATCCACCGTTTCTACGAATCTGGTATCGTATTCGAAGATAC 1206
|||||
403 LysAspTyrPheCysIleSerLysTyrTyrLeuLysGluValLysArgLy 419
|||||
1207 AAAGACTC.TTCTGCATCTCCAATGGTACTGAGGAAGTTAAACGCAA 1255
|||||
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProLysAspG 436
|||||
1256 ACGGTACAACCTGAAACTGGTTGCAATGGCAGTTTCATCCCGAAAGACG 1305
|||||
436 luGlyTrpThrGlu 440
|||||
1306 AAGGTTGGACCGAA 1319
|||||
```

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV30580

seq\_documentation\_block:

ID AAV30580 standard; DNA; 1526 BP.

XX AAV30580;

XX AC AAV30580;

XX DT 07-DEC-1998 (first entry)

XX DE Clostridium botulinum toxin B fragment C gene in pHisBotb.

XX KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;

XX KW botulism; BotB; ds.

XX OS Clostridium botulinum serotype B strain Eklund 17B.

XX OS Synthetic.

XX Key Location/Qualifiers

XX CDS 108..1526

XX FT /\*tag= a

XX PN WO9808540-A1.

XX PD 05-MAR-1998.

XX PF 28-AUG-1997; 97WO-US15394.

XX PR 28-AUG-1996; 96US-0704159.

XX PA (OPHI-) OPHIDIAN PHARM INC.

XX PI Thalley BS, Williams JA;

XX DR WPI; 1998-230234/20.

XX DR P-PSDB; AAW68393.

XX XX

PT Host cell containing recombinant expression vector encoding  
 PT Clostridium botulinum type B or E toxin - useful to treat humans  
 PT and other animals at risk of intoxication with clostridial toxin  
 XX  
 PS Example 35; Page 300-302; 428pp; English.

XX This is the DNA sequence of the Clostridium botulinum serotype B  
 CC toxin fragment C gene contained in plasmid pHsBotB. The encoded  
 CC fragment C polypeptide (see AAW68393) has a histidine-tagged  
 CC N-terminal extension. The vector was used to express native  
 CC (i.e. non-fusion) soluble C fragment in Escherichia coli host  
 CC cells. The invention relates to recombinant proteins derived from  
 CC C. botulinum toxins. Methods are provided which allow for the  
 CC isolation of soluble recombinant proteins free of significant  
 CC endotoxin contamination. Preferred hosts for production of  
 CC recombinant proteins are E. coli, insect cells and yeast cells.  
 CC The recombinant toxins are used as immunogens for the production  
 CC of vaccines and antitoxins that are useful in the treatment of  
 CC humans and animals at risk of intoxication with clostridial toxin.  
 XX

SQ Sequence 1526 BP; 622 A; 143 C; 259 G; 502 T; 0 other;

# ignment\_scores:

Quality: 2102.00 Length: 440  
 Ratio: 4.958 Gaps: 0  
 Percent Similarity: 96.364 Percent Identity: 88.636

## alignment\_block:

US-09-910-186A-8 x AAV30580

Align seg 1/1 to: AAV30580 from: 1 to: 1526

1 MetAlaAsnLysTyrAsnSerGluIleLeuAsnAsnIleLeuAsnIle 17  
 204 ATGTTTAAATAAATAGCGAAATTTTAAATAATATATCTTAAATTT 253  
 17 uArgTyrLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysV 34  
 254 AAGATATAGAGATAATAATTTAATAGATTATCAGGATATGGAGCAAGG 303  
 34 aGluValTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeu 50  
 304 TAGAGGTATATGATGGGTCAAGCTTAATGATATAAAATCAATTTAAATTA 353  
 51 ThrSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleI 67  
 354 ACTAGTTGACGACAGATAGTAAGTTAGAGTCACTCAAAATCAGAATATTAT 403  
 67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84  
 404 ATTTAATAGTATGTTCTTGTATTTTACGTTAGCTTTTGGATAAGGATAC 453  
 84 roLysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThr 100  
 454 CTAAATATAGGAATGATATACAAAATTTATTCATAATGAATATACG 503  
 101 IleIleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgG 117  
 504 ATAATTAATTTGATGAAATAAATTCAGCGTGGAAATATCTATTAGGGG 553  
 117 yAsnArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerV 134  
 554 TAATAGGATAATATGACCTTAATTTGATATAAATGGAATAAATCAATCAG 603  
 134 alPhePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArg 150  
 604 TATTTTTCATATATACATAGAGAAGATATATCAGAGTATATAAATAGA 653  
 151 TrpPhePheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrI 167  
 654 TGGTTTTTGTAACTATTACTAATAAATTTGGATAATGCTAAATTTATAT 703

167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184  
 704 TAATGGCAGCTTAGAATCAAAATATGGATATTAAGAGATATAGGAAGTTA 753  
 184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200  
 754 TTGTTAATGGTGAATAACATTTAAATTTAGATGGTGTAGATAGAGAACA 803  
 201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerG 217  
 804 CAATTTATTTGGATGAAATATTTTAGTATTTTAAATACCAATTAATACA 853  
 217 nSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeu 234  
 854 ATCAAAATTTAAAGAGATATATAAAATTTCAATCATATAGCGAATACTTAA 903  
 234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250  
 904 AAGATTTTGGGAAATCCTTTAATGTATAATAAAGAATATATATATGTTT 953  
 251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProVa 267  
 954 AATGGGGGAATAAAATTCATATATTAACTAGTGAAGATTTCATCTGT 1003  
 267 lGlyGluIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleA 284  
 1004 AGGTGAAATATTAATACGTAGCAAAATATAATCAGAAATTTCCAATTTATAA 1053  
 284 snTyrArgAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSer 300  
 1054 ATTATAGAAATTTATATATTTGGAGAAAAATTTATTTAAGAGAGAGTCA 1103  
 301 AsnSerGlnSerIleAsnAspAspIleValArgLysGluAspTyrIleTy 317  
 1104 AATCTCAATCTATAAATGATGATATAGTTAGAAAGAGAGATTATATACA 1153  
 317 rLeuAspPhePheAsnLeuAsnGlnGluTrpArgValTyrTyrLysT 334  
 1154 TCTAGATTTGGTACTTCACCATCAAGAGTGGAGAGTATATGCTATAAAT 1203  
 334 yrPheLysLysGluGluLysLeuPheLeuAlaProIleSerAspSer 350  
 1204 ATTTTAGGAACAGAGAAATAATTTGTTTATCTATATTAAGTGATTCT 1253  
 351 AspGluLeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProTh 367  
 1254 AATGNAATTTTATAGACTATAGAAATAAAGAAATATGATGACACCATC 1303  
 367 rTyrSerCysGlnLeuLeuPheLysLysAspGluSerThrAspGluI 384  
 1304 ATATAGTTGTCAGTTGCTTTTAAAAAAGATCAAGAAAGTACTGATGATA 1353  
 384 leGlyLeuIleGlyIleHisArgPheTyrGluSerGlyIleValPheGlu 400  
 1354 TAGGATGATTGGTATTCATCGTTTCTACGAATCTGGAGTTTACGTAAA 1403  
 401 GluTyrLysAspTyrPheCysIleSerLysTrpTyrLeuLysGluVal 417  
 1404 AAGTATAAAGATATTTTGTATAGTAATAGGTACTTTAAAGAGGTAAA 1453  
 417 sArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProL 434  
 1454 AAGAAACCATATAAGTCAAAATTTGGGATGTAATTTGGCAGTTTATTCTTA 1503  
 434 ysAspGluGlyTrpThrGlu 440  
 1504 AAGATGAAGGTGGACTGAA 1523

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV26289

seq\_documentation\_block:

ID AAV26289 standard; DNA; 3509 BP.

XX

AAV26289;  
 27-JUL-1998 (first entry)  
 Recombinant botulinum neurotoxin type B LH728/B encoding DNA.  
 Botulinum; recombinant; Clostridium botulinum; neurotoxin;  
 Immunogen; detection; tetanus; non-toxic; toxin; ds.  
 Synthetic.  
 OS Clostridium botulinum.  
 XX Key Location/Qualifiers  
 FH 1..3509  
 FT CDS /tag= a  
 FT /product= "LH728/B"  
 FT /note= "no stop codon given"  
 WO9807864-A1.  
 26-FEB-1998.  
 22-AUG-1997; 97WO-GB02273.  
 13-DEC-1996; 96GB-0025996.  
 23-AUG-1996; 96GB-0017671.  
 (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
 (SPEY-) SPEYWOOD LAB LTD.  
 Foster KA, Quinn CP, Shone CC;  
 WPI: 1998-169168/15.  
 P-PSDB; AAV56017.  
 Recombinant neurotoxin polypeptides - used to develop therapeutic agents, immunogens or as non-toxic standards for the detection of neurotoxins  
 Disclosure; Page 87-91; 137pp; English.  
 The present sequence encodes a recombinant neurotoxin protein from the present invention. The present invention describes recombinant neurotoxin proteins which comprise a first and second domain, where the first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis, and where the second domain is adapted: (a) to translocate the protein into a cell; (b) to increase the solubility of the protein compared to the solubility of the first domain on its own, or (c) both to translocate the protein into a cell and to increase the solubility of the protein compared to the solubility of the first domain on its own, the protein being free of clostridial neurotoxin (CN) and free of CN precursor that can be converted into toxin by proteolytic action. The recombinant proteins can be used as therapeutic agents for targeting cells expressing a relevant substrate. The products can also be used as immunogens and as non-toxic standards for the assessment and development of in vitro assays for the detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples.

SQ Sequence 3509 BP; 1468 A; 340 C; 548 G; 1153 T; 0 other;

# alignment\_scores:

Quality: 1667.00 Length: 318  
 Ratio: 5.259 Gaps: 0  
 Percent Similarity: 99.686 Percent Identity: 99.686

# alignment\_block:

US-09-910-186A-8 x AAV26289 ..

Align seg 1/1 to: AAV26289 from: 1 to: 3509

1 MetAlaAsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeu 17

|||||  
 2554 ATGTTTAAATAATATAGCGAAATTTTAAATAATATATATCTTAAATTT 2603  
 17 uArTYrLYsAspAsnAsnLeuIleAspLeuSerGlyTYrGlyAlaLysV 34  
 |||||  
 2604 AAGATATAAGGATATAATATTTTAAATAGATTTATCAGGATATGGGCAAGG 2653  
 34 aLGluValTYrAspGlyValGluLeuAsnAspLysAsnGlnPheLYsLeu 50  
 |||||  
 2654 TAGAGGTATATGATGGAGTCGAGCTTAATGATAAAATCAATTTAAATTA 2703  
 51 ThrSerSerAlaAsnSerLYsIleArgValThrGlnAsnGlnAsnIleIl 67  
 |||||  
 2704 ACTAGTTCAGCAATAGTAGATTAAGATTAGAGTGACTCAAAATCAGAATATCAT 2753  
 67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84  
 |||||  
 2754 ATTTAAATAGTGTCTTCCTTGATTTTAGCTTTAGCTTTTGGATAAGAATAC 2803  
 84 rOLysTYrLYsAsnAspGlyIleGlnAsnTYrIleHisAsnGluTYrThr 100  
 |||||  
 2804 CTAATATAAGAATAGTGGTATACAAATATATTCATAATGAATATATACA 2853  
 101 IleIleAsnCysMetLYsAsnAsnSerGlyTrpLYsIleSerIleArgGl 117  
 |||||  
 2854 ATAATTAATTCGTATGCAAAATAATTCGGCTGGAAATATATCTATTAGGGG 2903  
 117 yAsnArgIleIleTrpThrLeuIleAspIleAsnGlyLYsThrLYsSerV 134  
 |||||  
 2904 TAATAGGATAATATGGACTTTAATTTGATATAAATGGAATAACCAATCGG 2953  
 134 alPhePheGluTYrAsnIleArgGluAspIleSerGluTYrIleAsnArg 150  
 |||||  
 2954 TATTTTGAATAATAACATAAGAGAAGATATATCAGAGTATATAAATAGA 3003  
 151 TrpPhePheValThrIleThrAsnAsnLeuAsnAsnAlaLYsIleTYrIl 167  
 |||||  
 3004 TGGTTTTTGTACTATTACTAATAATTTGAATACGCTAAATTTATAT 3053  
 167 eAsnGlyLYsLeuGluSerAsnThrAspIleLYsAspIleArgGluValI 184  
 |||||  
 3054 TAATGGTAAGCTAGAAATCAAAATACAGATATTTAAAGATATAAGAGAAGTTA 3103  
 184 leAlaAsnGlyGluIleIlePheLYsLeuAspGlyAspIleAspArgThr 200  
 |||||  
 3104 TTGCTAATGGTGAATAATATTTAAATTAGATGGTGATATAGATAGAACA 3153  
 201 GlnPheIleTrpMetLYsTYrPheSerIlePheAsnThrGluLeuSerGl 217  
 |||||  
 3154 CAATTTATTTGGATGAAATATTTTCAGTATTTTAAACGGAATTAAGTCA 3203  
 217 nSerAsnIleGluGluArgTYrLYsIleGlnSerTYrSerGluTYrLeuL 234  
 |||||  
 3204 ATCAAAATATTGAAGAAGATATAAAATTCAAATCATATAGCGAATATTTAA 3253  
 234 ysAspPheTrpGlyAsnProLeuMetTYrAsnLYsGluTYrTYrMetPhe 250  
 |||||  
 3254 AAGATTTTGGGGAATACCTTTAATGTACAATAAAGAAATATATATATGTTT 3303  
 251 AsnAlaGlyAsnLYsAsnSerTYrIleLYsLeuLYsLYsAspSerProVa 267  
 |||||  
 3304 AATCGGGGAATAAAATTCATATATTTAAACTAAAGAAAGATTCCACCTGT 3353  
 267 lGlyGluIleLeuThrArgSerLYsTYrAsnGlnAsnSerLYsTYrIleA 284  
 |||||  
 3354 AGGTGAAATTTTAAACACGTAGCAAAATATAATCAAAATTTCTAAATATATAA 3403  
 284 snTYrArgAspLeuTYrIleGlyGluLYsPheIleIleArgLYsSer 300  
 |||||  
 3404 ATTATAGAGATTATATTTGGAGAAAAATTTATTTAAGAGAAAGATCA 3453  
 301 AsnSerGlnSerIleAsnAspIleValArgLYsGluAspTYrIleTY 317  
 |||||





```

67 ePheAsnSerValPheLeuAspPheSerValSerPheThrIleArgIleP 84
   :::::::::::::::::::::
240 CTAGGACTCCATGTTCCGACAACTTCTCCATCAACTTCTGGGTCGTACC 289
   :::::::::::::::::::::
84 roLysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThr 100
   :::::::::::::::::::::
290 CAAAGTACAAACAACACGACATCCAGACCTACCTCGACAGCAGTAGTACAC 339
   :::::::::::::::::::::
101 IleIleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgG 117
   :::::::::::::::::::::
340 ATCACTCTCTATCAAGACGACCTCCGGTTGGAAGGTTCTCCATCAAGG 389
   :::::::::::::::::::::
117 yAsnArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerV 134
   :::::::::::::::::::::
390 AAACCGTATCATCTGGACCTGATCGACGCTCAACGCCAAGTCCAAGTCCA 439
   :::::::::::::::::::::
134 aIlePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArg 150
   :::::::::::::::::::::
440 TCTTCTTCGAGTACTCCATCAAGGACAACTCTCGACTACATCAACAAG 489
   :::::::::::::::::::::
151 TrpPhePheValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTy 166
   :::::::::::::::::::::
490 TGGTTCCTCATCACCATCAACACGACCTCTGGGTAAACGCCAACATCTPA 539
   :::::::::::::::::::::
166 rIleAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluV 183
   :::::::::::::::::::::
540 CATCAACGGTTCCTCGAAGAACCTCCGAGAAGATCTGAACTGGACCGTGA 589
   :::::::::::::::::::::
183 aIleAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArg 199
   :::::::::::::::::::::
590 TCAACTCTCTCCAAACGACATPCGACTTCAAGCTGATCAACTGTACCGAC 639
   :::::::::::::::::::::
200 ThrGlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSe 216
   :::::::::::::::::::::
640 ACAAGTTCGCTGGATCAAGCACTTCACATCTTCGGTCTGGTGAAGTGA 689
   :::::::::::::::::::::
216 rGlnSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrL 233
   :::::::::::::::::::::
690 CGCCACCGAGGTCCTCTCCCTGCTACTGGATCCAGTCTCCACCAACACC 739
   :::::::::::::::::::::
233 euLysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTrMet 249
   :::::::::::::::::::::
740 TGAAGACTTCTGGGGAACCCACTCGCTTACGACACCCAGTACTACCTG 789
   :::::::::::::::::::::
250 PheAsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerPr 266
   :::::::::::::::::::::
790 TTCACACCGGTATCGACAGAACATCTACATCAAGTACTCTCCAAAGCCTC 839
   :::::::::::::::::::::
266 oValGlyGluIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrI 283
   :::::::::::::::::::::
840 CATGGGTGAGACCGCCCTCGTACCAACTTCAACAACGCCGCC.....A 883
   :::::::::::::::::::::
283 leAsnTyrArgAspLeuTyrIleGlyGluLysPheIleIleArgLys 299
   :::::::::::::::::::::
884 TCAACTACCAAGAACCTGTACCTGGGTCTGCGTTTCATCATCAAGAAGGCC 933
   :::::::::::::::::::::
300 SerAsnSerGlnSerIle...AsnAspIleValArgLysGluAspTyr 315
   :::::::::::::::::::::
934 TCCAACCTCCGTAACTACATCAACAACGACCAACATCTGCTCGTGGGGTGA 983
   :::::::::::::::::::::
315 rIleTyrLeuAspPhePheAsnLeuAsnGlnGlu...TrpArgValTyrT 331
   :::::::::::::::::::::
984 CATCTACCTGAACTACGACAAACATCTCCGACGAGTCTTACCGTGTCTAGC 1033
   :::::::::::::::::::::
331 hrTyrLysTyrPheLysLysGluGluLysLeuPheLeuAlaProIle 347
   :::::::::::::::::::::
1034 TCCCTGGTCAACTCCAAAGGAGATCCAGACCCAGCTGTCTCTGGCCCCAATC 1083
   :::::::::::::::::::::
348 SerAspSerAspLeuTyrAsnThrIleGlnIleLysGluTyrAspG 364
   :::::::::::::::::::::
1084 AACGACGACCTTACCTTCTACGACGCTCTGCGACATCAAGAAGTACTACGA 1133
   :::::::::::::::::::::
364 uGlnProThrTyrSerCysGlnLeuLeuPheLysLysAspGluGluSerT 381

```

```

1134 GAAGACCACTACAACCTGTCTAGATCTGTGCGAGAAGGAC.....A 1174
   :::::::::::::::::::::
381 hrAspGluIleGlyLeuIleGlyIleHisArgPheTyrGluSer...Gly 396
   :::::::::::::::::::::
1175 CCAAGACCTTCGGACTGTTCGGTATCGGTAACTTCGTCAAGGACTACGGT 1224
   :::::::::::::::::::::
397 IleValPheGluGluTyrLysAspTyrPheCysIleSerLysTrpTyrIle 413
   :::::::::::::::::::::
1225 TACGTCTGGGACACCTACGACAACTACTCTGTATCTCCAGTGGTACCT 1274
   :::::::::::::::::::::
413 uLysGluValLysArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpG 430
   :::::::::::::::::::::
1275 GGTCTGTATCTCCGAGAACATCAACAAGCTGGTCTGGATGTAACCTGCG 1324
   :::::::::::::::::::::
430 InPheIleProLysAspGluGlyTrpThrGlu 440
   :::::::::::::::::::::
1325 AGTTCACTCCAGTCGACGAGGGTTGGACCGAG 1356
   :::::::::::::::::::::
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30596
seq_documentation_block:
ID   AAV30596 standard; DNA; 1535 BP.
XX
AC   AAV30596;
XX
DT   07-DEC-1998 (first entry)
XX
DE   Clostridium botulinum type G toxin C fragment gene in pETHsb.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin G; intoxication; immunogen;
    botulism; BotG; ds.
XX
OS   Clostridium botulinum serotype G 113/30 strain.
XX
FH   Key Location/Qualifiers
FT   CDS 108..1529
FT     /*tag= a
XX
PN   W09808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997; 97WO-US15394.
XX
PR   28-AUG-1996; 96US-0704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Thalley BS, Williams JA;
XX
DR   WPI; 1998-230234/20.
DR   P-PSDB; AAW68400.
XX
PT   Host cell containing recombinant expression vector encoding
    Clostridium botulinum type B or E toxin - useful to treat humans
    and other animals at risk of intoxication with clostridial toxin
XX
PS   Example 49; Page 376-378; 428pp; English.
XX
CC   This is the DNA sequence of the Clostridium botulinum serotype G
    (113/30 strain) neurotoxin fragment C gene contained in plasmid
    pETHsb. The encoded BotG fragment C polypeptide (see AAW68400) has a
    His-tagged N-terminal extension. The vector can be used to express
    native (i.e. non-fusion) soluble C fragment in Escherichia coli host
    cells. The invention relates to recombinant proteins derived from
    C. botulinum toxins, especially type B and type E toxins. Methods
    are provided which allow for the isolation of soluble recombinant
    proteins free of significant endotoxin contamination. Preferred
    hosts for production of recombinant proteins are E. coli, insect
    cells and yeast cells. The recombinant toxins are used as
    immunogens for the production of vaccines and antitoxins that are

```



CC useful in the treatment of humans and animals at risk of  
CC intoxication with clostridial toxin.  
CC  
XX  
SQ Sequence 1535 BP; 596 A; 172 C; 246 G; 521 T; 0 other;

```
alignment_scores:
  Quality: 1168.50
  Ratio: 3.246
  Percent Similarity: 81.081
  Percent Identity: 50.000
  Length: 444
  Gaps: 7
```

```
alignment_block:
US-09-910-186A-8 x AAV30596
```

Align seq 1/1 to: AAV30596 from: 1 to: 1535

3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleLeuAsnLeuArgTy 19  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
210 AATAATATATATAGTAATATAGTAGTACGTATTATTTAAGTTAAAGTAA 259

19 rLyAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
260 TAGAGGTGGCGTTTATAGATTCATCTGGATATGGTGCAACTATGAATG 309

36 alTyAspGlyValGluLeuAsnAsp . . . . . LysAsnGlnPheLysLeu 50  
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||  
310 TAGGTTCCAGATGGTTATCTTTAATCATATAGCAAAATGCTCAATTTAAATTA 359

51 ThrSerSerAlaAsnSerLysIleArgValThrClnAsnClnAsnIleIl 67  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
360 AATAATCTTGAAATAGTATATATACGGCACATCAAAGTAAATTCGTTGT 409

67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
410 ATATGATAGTATGTTTGATAATTTAGCAATTAACCTTTGGGTAAAGCACTC 459

84 roLysTyrLysAsnAspGlyIleClnAsnTyrIleHisAsnGluTyrThr 100  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
460 CTAATATAATAATAATGATATACAACCTATCTCCAATGAGTAGTACA 509

101 IleIleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGl 117  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
510 ATAATTAGTGTATAAAAATGACTCAGGATGGAAAGTATCTATTAAAGG 559

117 yAsnArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysServ 134  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
560 AAATAGAATAATATGGACATTAATAGATGTTAATGC AAAATCTAAAAATCAA 609

134 alPhePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArg 150  
:||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
610 TATTTTTCGAATATAGTATAAAGATAAATATATCAGATTATATAATAAA 659

151 TrpPhePheValThrIleThrAsnAsn . . . LeuAsnAsnAlalysIleTy 166  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
660 TGGTTTTCCATAACTATTACTAATGATAGATTAGGTAAACGCAAAATATTTA 709

166 rIleAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluV 183  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
710 TATAAATGGAAAGTTGAAAAAAGTGA AAAAATTTTAAACTTAGATAGAA 759

183 alIleAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArg 199  
||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
760 TTAATTTCTAGTAATGATATAGACTTCAATTAATTAATTTGTACAGATACT 809

200 ThrGlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSe 216  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
810 ACTAAATTTGTTTGGATTAAGGATTTTAATATTTTGGTAGACAATTA 859

216 rGlnSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrL 233  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
860 TGCTACAGAAGATPTCTTCATATATGGATTCATCATCTACAAATACTPT 909

233 euLysAspPheTrpGlyAsnProLeuMetTyrAsnLvsGluTyrTyrMet 249

|||||  
910 TAAAGATTTTGGGGAATCCTTTAAAGATACGATACACAACTACTATCTG 959  
250 PheAsnAlaGluAsnLysAsnSerTyrTrileysLeuLysLysAspSer 266  
|||||  
960 TTTAATCAAGGTATGCAAAATATCTATATAAAGTATTTTAGTAAGCTTC 1009  
266 ovalGluLeuLeuThrArgSerLysTyrAsnGlnAsnSerLysTyr 283  
|||||  
1010 TATGGGGGAAACTGCACCACCTGACAACTTTAAATAATGCGACGA. . . .A 1053  
283 leAsnTyrArgAspLeuTyrileGlyGluLysPheilelleArgAtgLys 299  
|||||  
1054 TAAATATCAAAATTTATATCTTGTTTACGATTTATATAAAAAGCA 1103  
300 SerAsnSerGlnSerile. . .AsnAspAspIleValArgLysGluAspTyr 315  
|||||  
1104 TCAAAATCTCGGAATATAAAATAATGATAATATAGTCAGAGAAGGAGATTA 1153  
315 rIleTyrTrileuAspPhePheAsnLeuAsnGlnGlu. . .TrpArgValTyr 331  
|||||  
1154 TATATATCTTAATATGATAATATCTTGATGAATCTTACAGAGATATATG 1203  
331 hrTyrLysTyrPhelysLysGluGluGluLysLeuPheLeuAlaProile 347  
:: ||||::  
1204 TTTTGGTGAATCTAAAGAAATTCAAACTCAATTTATTTTAGCACCCATA 1253  
348 SerAspSerAspGluLeuTyrAsnThrIleGlnileLysGluTyrAspG 364  
:: ||||::  
1254 AATGATGATCTCAGCTTCTATGATGCTACTACAAATAAAAAATATTATGA 1303  
364 uGlnProThrTyrSerCysGlnLeuLeuPheLysLysAspGluGluSer 381  
1304 AAAACAACATATAATGTCAGATACCTTGGCAAAAAGAT. . . . .A 1344  
381 hrAspGluileLysLeuileGlyIleHisArgPheTyrGluSer. . .Gly 396  
|||||  
1345 CTAAACAATTTGGGCTGTTCGAATTTGGTAAATTTGTTAAAGATTATGGA 1394  
397 lleValPheGluTyrLysAspTyrPheCysIleSerLysTyrPyrLe 413  
|||||  
1395 TATGTTTGGGATACCTATGATAATATTTTGGCAATAAGTCAGTGGTATCT 1444  
413 uLysGluValLysArgLysProTyrAsnLeuLeuGlyCysAsnTrpG 430  
|||||  
1445 CAGAAGATATCTGAAAATATAAATAATTAAGGTTGGGATGTAATTGGC 1494  
430 lnPheileProLysAspGluGlyTrpThrGlu 440  
1495 AATTCATTTCCGTGGATCGAAGGATGGACAGAA 1526

seq name: /SIPSI/qcadata/hold-geneseq/geneseqn-emb1/NA2000 DAT:AAZ87220

seq documentation block:

seq\_documentation\_block:  
ID AAZ87220 standard: DNA: 1317 BP.

AA  
AC AA787220:XX  
XX  
1077/0777

08-MAY-2000 (first entry)

DE DNA encoding native BoNT serotype A (BoNT<sup>SA</sup>) C-terminal fragment (Hc).

KW Botulinum neurotoxin: heavy chain: BoNT: serotype A: XX

KW Botulinum neurotoxin; heavy chain; BONT; serotype A;  
KW C-terminal fragment; Hc: Venezuelan equine encephalitis virus replicon;

KW VEE: botulism: vaccine: diagnosis: drug screening: ds. c-terminal fragment; HC, Venezuelan equine encephalitis

XX Clostridium botulinum

XX	Key	Location/Qualifiers
EH		

	FH	Key
FT		CDC

FT FT CDS

JF JF

33

### ET



```
358 nileLysGluTyrAspGluGlnPro...ThrTyrSerCysGlnLeuLeuP 374
:::||||: ::::||||| ||| ::::||||:
1092 AATGAAGTCAAAAATGATCAAGGAATAACAAATAATGCAAAATG... 1137
:::||||: ::::||||| ||| ::::||||:
374 helLysLysAspGluSerThrAspGluLeuLeuGlyIleHis 390
:::||||: ::::||||| ||| ::::||||:
1138 ..AATTACAAAGATAAATGGAATGATATAGGCTTTATAGATTTCAT 1185
:::||||: ::::||||| ||| ::::||||:
391 ArgPheTyrGluSerGlyIleValPheGluGluTyrLysAspTyrPheCy 407
:::||||: ::::||||| ||| ::::||||:
1186 CAGTTTAATAATATAGCT.....AAACTAGT 1211
:::||||: ::::||||| ||| ::::||||:
407 sileSerLysTrpTyrLeuLysGluValLysArgLysProTyrAsnLeuL 424
:::||||: ::::||||| ||| ::::||||:
1212 AGCAAGTAATGTGTAATAAGACAAATAGAAAGATCT.....AGTAGGA 1255
:::||||: ::::||||| ||| ::::||||:
424 ysLeuGlyCysAsnTrpGlnPheIleProLysAspGluGlyTrpThrGlu 440
|||||:||||: ::::||||| ||| ::::||||:
1256 CTTTGGGTTGCTCATGGGAATTTATTCCTGTAGATGATGGGAGAA 1305
|||||:||||: ::::||||| ||| ::::||||:
```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAA54483

seq\_documentation\_block:

ID AAA54483 standard; DNA; 1323 BP.

AC AAA54483;

DT 11-APR-2001 (first entry)

DE Botulinum toxin heavy chain C-terminal coding sequence (serotype A).

KW Botulinum toxin; neurotoxin; heavy chain; recombinant expression;

KW recombinant vector; antigen; immune response; vaccine; bacterium;

KW infection; ds.

OS Synthetic.

OS Clostridium botulinum.

FH Key Location/Qualifiers

FT CDS 13..1317

FT /\*tag= a

FT /product= H\_C peptide fragment

XX WO200067700-A2.

XX 16-NOV-2000.

XX 12-MAY-2000; 2000WO-US12890.

XX 12-MAY-1999; 99US-0133865.

XX 12-MAY-1999; 99US-0133866.

XX 12-MAY-1999; 99US-0133867.

XX 12-MAY-1999; 99US-0133868.

XX 12-MAY-1999; 99US-0133869.

XX 12-MAY-1999; 99US-0133873.

XX 29-JUL-1999; 99US-0146192.

XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.

XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;

XX WPI; 2001-016048/02.

XX P-PSDB; AAB04089.

XX New nucleic acids encoding the carboxy- or amino-terminal portions of  
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
PT vaccine against botulism

XX Disclosure; Fig 2a; 73pp; English.

XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide

CC chain and then posttranslationally nicked, forming a dichain

CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
CC remain linked by a disulfide bond. Nucleic acids encoding the  
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
CC expression vectors and expressed in transformed cells to produce  
CC peptide antigens useful for eliciting an immune response to give  
CC protective immunity against botulinum neurotoxin, which causes  
CC botulism. The nucleic acids are expressible in a recombinant  
CC organisms such as Escherichia coli or Pichia pastoris. The use  
CC of recombinant nucleic acids are advantageous since it eliminates  
CC the need to culture large quantities of hazardous toxin-producing  
CC bacterium. Production yield from the genetically engineered product  
CC is also high and cost of production is lower. The nucleic acids can  
CC be derived from Clostridium botulinum serotypes A-G.

XX  
SQ Sequence 1323 BP; 404 A; 334 C; 242 G; 343 T; 0 other;

alignment\_scores:

Quality: 831.50 Length: 450

Ratio: 2.607 Gaps: 10

Percent Similarity: 70.889 Percent Identity: 40.000

alignment\_block:

US-09-910-186A-8 x AAA54483 ..

Align seg 1/1 to: AAA54483 from: 1 to: 1323

4 LysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTyrLy 20

:::||||: ::::||||| ||| ::::||||:

28 GAATACATCAAGAATCATCATACCTCCATCCTGAACCTGCGCTACGA 77

:::||||: ::::||||| ||| ::::||||:

20 sAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluVal 37

:::||||: ::::||||| ||| ::::||||:

78 ATCCAACTACCTGACCTGCTCGCTACGCTTCCAAATCAACATCG 127

:::||||: ::::||||| ||| ::::||||:

37 yTyrAspGlyValGluLeuAsn.....AspLysAsnGlnPheLysLeuThr 51

:::||||: ::::||||| ||| ::::||||:

128 GTTCTAAAGTTTAACTTCGATCCGATCGACAAAGAAATCAGATCCAGCTGTC 177

:::||||: ::::||||| ||| ::::||||:

52 SerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePh 68

:::||||: ::::||||| ||| ::::||||:

178 AATCTGGAATCTTCCAAAATCGAAGTTATCTCTGAAGAATGCTATCGTATA 227

:::||||: ::::||||| ||| ::::||||:

68 eAsnSerValPheLeuAspPheSerValSerPheTyrIleArgIleProL 85

:::||||: ::::||||| ||| ::::||||:

228 CAACTCTATGTACGAAACCTCTCCACCTCTTCTGGATCCGATCCCGA 277

:::||||: ::::||||| ||| ::::||||:

85 yTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIle 101

:::||||: ::::||||| ||| ::::||||:

278 AATACTTCAAC.....TCCATCTCTGAACAATGAATACACCATC 318

:::||||: ::::||||| ||| ::::||||:

102 IleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAs 118

:::||||: ::::||||| ||| ::::||||:

319 ATCAACTGCTGGAACAACTCTGCTGTGGAAGTATCTCTGAACATCGG 368

:::||||: ::::||||| ||| ::::||||:

118 nArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValP 135

:::||||: ::::||||| ||| ::::||||:

369 TGAATCATCTGGACTCTCGAGGACACTCAGGAAATCAAAACGCGTGTG 418

:::||||: ::::||||| ||| ::::||||:

135 hePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrp 151

:::||||: ::::||||| ||| ::::||||:

419 TATTCAATACTCTCAGATCATCAACATCTCTGACTACATCAATCGCTGG 468

:::||||: ::::||||| ||| ::::||||:

152 PhePheValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrI 167

:::||||: ::::||||| ||| ::::||||:

469 ATCTTCGTTACCATCAACCAATCTGCTGTAATAACTCCAAAAATCTACAT 518

:::||||: ::::||||| ||| ::::||||:

167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluVal 184

:::||||: ::::||||| ||| ::::||||:

519 CAACGCGCGCTGATCGACCAACCGATCTCCAAATCTGGGTAACATCC 568

:::||||: ::::||||| ||| ::::||||:

184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200



PT Fusion proteins comprising non-toxin protein and part of toxin -  
PT useful to form anti-toxins against Clostridium botulinum type A, and





1042 TTGTCGTCTCGAAATCCCGACGTTGGTAATCTGTCTCAGGTAGTTCT 1091  
358 nileLysGluTyrAspGluInPro...ThrTyrSerCysGlnLeuLeup 374  
1092 AATGAATCCAGACGACCGAGGTATCTACTAACAAATGCAAAATG.... 1137  
374 helLysLysAspGluGluSerThrAspGluIleGlyLeuIleHis 390  
1138 ...AATCTGCAGCACACAAATGGTAACGATATCGGTTTCATCGGTTTCCAC 1185  
391 ArgPheTyrGluSerGlyIleValPheGluGluTyrLysAspTyrPheCy 407  
1186 CAGTTCACAAATATCGCT.....AAACTGCT 1211  
407 sileSerLysTyrTyrLeuLysGluValLysArgLysProTyrAsnLeuL 424  
1212 TGCCTTCCAACTGGTACAAATCGTCAGATCGACGTTCC.....TCTCGCA 1255  
424 ysLeuGlyCysAsnTrpGlnPheIleProLysAspGluGlyTrpThrGlu 440  
1256 CTTCTGGGTTGCTCTGGGAGTTTCATCCCGGTTGATGACGCGTTGGGGTGAA 1305



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2002, 16:00:43 ; Search time 1859.46 Seconds  
(without alignments)  
15091.739 Million cell updates/sec

Title: US-09-910-186A-7

Perfect score: 1341

Sequence: 1 gaattcacgagtgccacaa.....ggaccgaatagtaagaattc 1341

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vi:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_ju:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vi:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

1	742.8	55.4	4041	1	CLOBOTB	M81186 Clostridium
2	706.4	52.7	3876	1	AF295926	AF295926 Clostridi
3	706.4	52.7	3876	1	AF300465	AF300465 Clostridi
4	706.4	52.7	3876	1	AF300466	AF300466 Clostridi
5	706.4	52.7	11170	1	CBY13630	Y13630 Clostridium
6	696	51.9	3869	1	AF300467	AF300467 Clostridi
7	696	51.9	3869	1	AF300468	AF300468 Clostridi
8	696	51.9	3869	1	AF300469	AF300469 Clostridi
9	681.4	50.8	1326	1	CB0242628	AJ242628 Clostridi
10	663.4	49.5	4051	1	CB0242628	AJ242628 Clostridi
11	522.4	39.0	3509	6	A69701	A69701 Sequence 19
12	522.4	39.0	3509	6	BD009886	BD009886 Recombina
13	341	25.4	6862	1	CBNTNHTB	X87849 C.botulinum
14	263	19.6	1330	6	AR000029	AR000029 Sequence
15	263	19.6	1330	6	AR169140	AR169140 Sequence
16	263	19.6	1330	6	AX036243	AX036243 Sequence
17	263	19.6	1338	12	XX022962	U22962 Synthetic b
18	263	19.6	1402	6	AR000030	AR000030 Sequence
19	263	19.6	1402	6	AR169141	AR169141 Sequence
20	263	19.6	1402	6	AX036246	AX036246 Sequence
21	242.6	18.1	1313	6	A58946	A58946 Sequence 6
22	241	18.0	1299	12	AF251281	AF251281 Synthetic
23	224.8	16.8	1084	1	CBPPTOT	X70817 C.botulinum
24	220.2	16.4	3937	1	CBPPTOT	X74162 C.botulinum
25	212.6	15.9	1084	1	CBPPTOT	X70814 C.botulinum
26	212.6	15.9	1084	1	CBPPTOT	X70819 C.botulinum
27	135.4	10.1	4067	1	CBNTQXA	X73423 C.botulinum
28	134.2	10.0	3712	6	A49987	A49987 Sequence 4
29	131.4	9.8	1359	6	I28431	I28431 Sequence 3
30	129.4	9.6	4366	6	A42484	A42484 Sequence 12
31	129.2	9.6	3754	6	A37074	A37074 Sequence 17
32	129.2	9.6	3754	6	A42478	A42478 Sequence 6
33	129.2	9.6	3769	6	A37075	A37075 Sequence 18
34	129.2	9.6	3769	6	A49988	A49988 Sequence 5
35	129.2	9.6	3769	6	A42481	A42481 Sequence 9
36	123	9.2	3891	6	AR000031	AR000031 Sequence
37	123	9.2	3891	6	AR169142	AR169142 Sequence
38	123	9.2	3891	6	AX036248	AX036248 Sequence
39	123	9.2	4292	1	CBOTAG	X52066 Clostridium
40	123	9.2	4835	1	CLONEUR	M30196 C.botulinum
41	103.2	7.7	1766	6	A49889	A49889 Sequence 6
42	96	7.2	1293	6	A58945	A58945 Sequence 5
43	96	7.2	4199	1	CLOBONT	I35496 Clostridium
44	96	7.2	4209	1	CBPPTOT	X81714 C.botulinum
45	92.6	6.9	3952	1	CBNTTD	X54254 Clostridium

#### ALIGNMENTS

RESULT	1	CLOBOTB	Clostridium botulinum neurotoxin type B (botB) gene, complete cds.
LOCUS	1	CLOBOTB	4041 bp DNA linear BCT 26-APR-1993
DEFINITION	1	Clostridium botulinum neurotoxin type B (botB) gene, complete cds.	
ACCESSION	1	M81186	
VERSION	1	M81186.1	GI:144734
KEYWORDS	1	botB gene; neurotoxin type B.	
SOURCE	1	Clostridium botulinum DNA.	
ORGANISM	1	Clostridium botulinum	
REFERENCE	1	Whelan, S.M., Elmore, M.J., Bodsworth, N.J., Brehm, J.K., Atkinson, T.	
AUTHORS	1	Whelan, S.M., Elmore, M.J., Bodsworth, N.J., Brehm, J.K., Atkinson, T.	
TITLE	1	Complete nucleotide sequence of the Clostridium botulinum gene	
JOURNAL	1	encoding the type B neurotoxin	
FEATURES	1	Unpublished (1991)	
source	1	Location/Qualifiers	
gene	1	1..4041	
	1	/organism="Clostridium botulinum"	
	1	/db_xref="taxon:1491"	
	1	57..3932	
	1	/gene="botB"	

CDS	57...3932				
	/gene="botB"				
	/function="vertebrate neurotoxin"				
	/codon_start=1				
	/transl_table=11				
	/product="neurotoxin type B"				
	/protein_id="AA23211.1"				
	/db_xref="GI:144735"				
	/translation="MPVTINNFNYNDPIDNNNIIMPEPPFARGTGYRYKAFKIDRIW IPERYTGYKPEDNKSSGIFNRDCEVYDDPDLNTNDKKNIFLOTWKLFNRLKSK PLGKLEMLINGIPYLGDRVPLEEFNTASVTYKLIISNPGEVERKKGIFANLII FGFSPVLENETIDIGIQNHFAASREGFGIMQKFCPEYVSFNVNQKNKAASIFNRR GYFDPALIMHDLIHVGLYKIVDDLPVPNKCFQPMQSTDAIQAEALYFGQD PSIITPDKSIIYDKVLQFRGIVDLNKLIVICSDPININIYKNNKFKDKYKFVEDS EKGYSIDVESDKLYKSLMFGFTETNIAENYKIKTRASYSDSIPVPIKLLDNEIV TEEGFNISDDMEYRGONKAINQAVEEISEKHLAVYKIQMCKSVKAPGICIDVD NEDLFFIADKNSFSDLSKNERIENYQSTNIENDFPINELLIDLDLSKIELPSENT ESLIDFNVDPVYEQAPAIKKIFDENTIFQYLSYTFPDLRIDISLSSFDLALLS NKVVSFFSMDYIKTANKVEAGLFAWYKQIVNDFVIBANKSNTMDKTDADSLIVPYI GLIANGVETAKGFENAFETAGASILLFPELLIPVGFALLESYIDNKNKIYTI DNALTKRNEKSDMYGLIVAOWLSVTQFTYIKEGMYKALNYOQAQLEELIKRYNI YSEKESNINIDFNDSKLNENQADINNFINGCSVYLMKKMPLAVEKLDF DNTLKNLLNIDENKLYLIGSAEYKSKVNYLKTIMPDLISITNTDILIEFNKY NSEILNILLNRKDNLLDLSYGAKVEYDVGELDNQKFLITSSANSKIRVTQN QNIIFSVELDFSVSFWIRIPKYKNDGQYI THNEYTI INCMKNNSWKISIRGNRII WTLIDINGTKTSFVEYNIREDISYINRWFVTITNNLNNAKIYINGKLESNTDIDK IREVTANGELIKFLDGDIDRTQFIWMKYVSIPTFELSSNIEERYKIQSYEYLKDFW GNPLMYKNEYMFNAGNKNYSIKLKQSPVGEILLTRSKYQNSKYNINRYDLYIGKFI IRKNSNSINDDIVRKEDYIYLDFFNLNQWRVYTYKFRKEEKLFLAPISDSDEF INTQIQEYDSOPTYSCOLLFKDEESTDEIGLIGHRFESGIVFEYKDYFCISKW YLKEYKRPYNLKLGCMNQFIPKDEGWE"				
BASE COUNT	1679 a	383 c	645 g	1334 t	
ORIGIN					
Query Match	55.4%;	Score	742.8;	DB 1;	Length 4041;
Best Local Similarity	72.4%;	Pred.	No. 1.1e-170;		
Matches	963;	Conservative	0;	Mismatches	367;
				Indels	0;
				Gaps	0;
QY	10	atggccaacaatacaattccgaattccggaacataatcctgaacacgcggttacaaa	69		
Db	2610	ATGTTTAAATTAATAATACGCAAAATTTTAAATAATATATATTTAAATTTAAGATATATAAG	2669		
QY	70	gacaaactatgatcgtatcgtctggttacggtgcgtcaagtgaagtacagcgggttt	129		
Db	2670	GATAATTAATTTATAGATTATACAGATATATGCGGCAAGGTAGAGGTATATGATGGAGTC	2729		
QY	130	gaactgaatgacaagaccaggttcaactgcacctcttcgcgttaactcctaagatccggttt	189		
Db	2730	GAGCTTAATGATATAAATCAATTTAAATTTAACTAGTTCAGCAAAATAGTAAGATTAGAGTG	2789		
QY	190	actcagaatcagacatcattctcaactccgtattcctcggaactctctgttccctctcg	249		
Db	2790	ACTCAAAATCAGAATATCATATTTAATAGTGTGTTCTTGTATTTTAGCGTTAGCTTTTGG	2849		
QY	250	attcgtatcccgaaaacagaacgacggtatccagaattatcatccacaatgaatacacacc	309		
Db	2850	ATAGAATACCTTAATATAGAATGATGTTATACAAATTTATATTCATATGAATATACATA	2909		
QY	310	atcatcaactgcataagaataactctggttggagaagatctccatccgcgcggttaacccgtatc	369		
Db	2910	ATAATTAATTTGATGAAAAATAATTCGGGCTGGAATAATCTATTAGGGGTAAATAGGATA	2969		
QY	370	attcgactctgatcatatcaacggtgaagaccaaactctgtattcttcctgaatacaaatc	429		
Db	2970	ATATGGACTTTAATGATATAATGAAAAACCAATCCGGTATTTTTCGATAATAACATA	3029		
QY	430	cgtgaagacatctctgaatacatcaactcgtcgtgttcttcgttaccatcaccaataaacctg	489		
Db	3030	AGAGAAGATATATCAGAGTATATAATAGATCGTTTTTGTAACTATTACTATAAATTTG	3089		
QY	490	aacactgctaaatctcacatcaacggtgaactgaacttaactcacgcagatcaatcaagacatc	549		





Qy 1276 ctgaaactgggttgcattggcagttccatcccaaaagacgaaggttgaccgaata 1331  
 ||| |||| ||| |||| |||| ||| ||| |||| |||| |||| |||| |||| |||| ||||  
 Db 3820 TCAAAATGGGATGTAATGGCAGCTTTATTCTTAAAGATGAAGGGTGACTGAATA 3875

RESULT 4  
 AF300466  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Clostridium botulinum isolate 593 type B cryptic neurotoxin gene, complete cds.  
 AF300466  
 AF300466.1 GI:15982938  
 Clostridium botulinum.  
 Clostridium botulinum  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 Clostridium

1 (bases 1 to 3876)  
 Kirma,N., Ferreira,J.L. and Baumstark,B.R.  
 Characterization of six type A strains of Clostridium botulinum that contain type B toxin gene sequences  
 Unpublished  
 2 (bases 1 to 3876)  
 Kirma,N., Ferreira,J.L. and Baumstark,B.R.  
 Direct Submission  
 Submitted (28-AUG-2000) Department of Biology, Georgia State University, P.O. Box 4010, Atlanta, GA 30302-4010, USA  
 Location/Qualifiers  
 1..3876  
 /organism="Clostridium botulinum"  
 /isolate="593"  
 /db\_xref="taxon:1491"  
 /country="USA: Georgia"  
 /note="isolated from dog feces in 1976"  
 type: A(B)  
 1..3876  
 /codon\_start=1  
 /transl\_table=1  
 /product="type B cryptic neurotoxin"  
 /protein\_id="AAL11499.1"  
 /db\_xref="GI:15982939"  
 /translation="MPVTINFNNDPDIINNIMPEPPFARGMGYYKAFKITDRIM  
 IIPERTFYGKDPFNKSGINRDVCEYDIDYLDNNDKNIQLTQMILKFNRTKRS  
 PLGKLEMLINIGIPYLDGRVPLDEFNTIASVTVNKLISNPFGEVRKKGIFANIL  
 FGPGVPLENETIDIGIQNHFAISREGFGIMQKFCPEYVSFNVNVOENKQASIFNR  
 GYSPDALILMELHVLHGLGVKIVNDLPIVPNEKKFEMQSTDAQELATFGGSD  
 PSLISPTDKSIYKVLQNPGRGIVRLKVLVCISDPNININIKNPKDKYKFFVEDS  
 EKGISIDVSEFKLYKSLMGFTETNAENYKIKTASFYSDSPVVKILNLDNIEY  
 TIEEGNIDKNMEYQONKAINKQAYEEISKEHLAVYKIQMCKSVKAPGICIDV  
 NEDLTFADKNPVSFSDLSKNERIAYTNNYIENDFSINELIDTLDISLTSSFDALLFS  
 ESILDFNVVYVYKQPAIKITIDENTIFQVLYSOTFFLDIRDLSLTSSFDALLFS  
 NKVYSPFSDMYLTKANKVYEVAGLFGAGVKQIVNDFVIEANKSSTMDKIDISLIVPI  
 GUALVNGNETAKGFENAEIAGASILLTEFIPELLIPVVGAFLESYIDNKKIETFI  
 NSALTRDKEDIMYGLIYAQMLSTVNTQFTYIKEGYKALNYQAALBEILIKYKNI  
 YSEKERSNINIDFNVDNKLNGINOAIDNINNFINECSYILMKKMIPLAVEKLLDF  
 DNTLRKNLLNIDENKYLIGSAEYKSKDYKLTSPIDFLSTYTNNTLILIEFNKY  
 NNDILNINILNRYRNDLDSGKGVYEDGVKLDNKQOFKLTSSANSKIRVTQNI  
 ONILNPSMFLRSYFWIRIPKYNKDQIONYTHNEVTIINCKNNSEWKISIRGNMII  
 WFLIDNGIKIKVFFEYSIKEDISEYINRWFFVTINNSDNAKIYINGKLESHDIRD  
 IREVIANDEIIFPKLDGNIDRTQFIWMKYPSEINTELQSNIIEIYKIQSYSEYLDKIF  
 GNPLMYNKYYFMNAGNKNSYIKLKDDSVGEILTRSKYQNQNSKINYRDLXIGEKFI  
 IRRKSNQSIDNDIVKEDYIYLDFFNLNQEWRYVMYKFKKKEEKLFLAPISDSDF  
 YNTIQIKDEYDEOPTYSCOLLFKDRESTDIEIGLIHIFVESGIVPKDYKDFCISK  
 YLKEYKRKPYNSKLGQWQFIPKDEGTE"

BASE COUNT 1611 a 367 c 616 g 1282 t  
 ORIGIN

Query Match 52.7%; Score 706.4; DB 1; Length 3876;  
 Best Local Similarity 71.0%; Pred. No. 8.8e-162;  
 Matches 935; Conservative 0; Mismatches 381; Indels 0; Gaps 0;

Qy 16 acaaaatacaattccgaatacttgaaacaatatcatctgaacctgcgttacaagaacac 75











Qy	76	aatctgatogatctgtctctgtgttacggtgtctaagttgaagtatacagacggtgttgtaactg	135
Db	2614	AAAGTTAATAGATTATACAGGATATGGGCAAGGTAGAGGTATATGATGGGGTCAAGCTT	2673
Qy	136	aatgacaagaaccagttccaactgacctcttcggtctaactctaagataccggtgttaactcag	195
Db	2674	AATGATAAAAATCAATTTAAATTAATCTAGTTCAGCAAAATAGTAAGATTAGAGTGCATCAA	2733
Qy	196	aatcagaacaatcatcttcaactccggtatctctcgacttctctgttctctctcgattcgt	255
Db	2734	AATCAGANATCATATTTAATAGTAGTGTCTTGATTTACGGTTAGTTTTCGATAAAGA	2793
Qy	256	atcccgaatacaagaacgacggtatccagaaattatcatccacaataaatacacocatcatc	315
Db	2794	ATACCTAAATATAAGAAATGATGGTATACAAAATTATATTCATATAATGAATATACAAATAAT	2853
Qy	316	aactgataaagaataactctggttgggaagatctccatccgcggtgaaccgtatactatcgtg	375
Db	2854	AANTGTATGAANAATAAATCTTGATGGGA-ATATCTATTAGGGGTAAATGATTAATATGG	2912
Qy	376	actctgatcgatatcaacggtgaagccaaactctgtattcttcgatacaacaatccgtgaa	435
Db	2913	ACTTTAATTGATATAAATCGAAAAATCAAAATCAGTATTTTTTGAATATAGCATAAAGAA	2972
Qy	436	gacatctctgaatacaatacaactcgtctgttcttcgttccatctcaacataaactcgaacaat	495
Db	2973	GATATATCAGAGTATATAAATAGATGGTTTTTGTGAACATTACTAATAATTCGGATTAAC	3032
Qy	496	gctaaaactctacataacgcgttaaacctgaaacttaataccgacatcaagaacatccgtgaa	555
Db	3033	GCTPAAATTTATTTAATCGTAACTAGTAAACATATAGATATATTAGAGATATAAGAGAA	3092
Qy	556	gttatcgctaaacggtgaaatcatcttcaaacctggacggtgacatcgatcgatccacgattc	615
Db	3093	GTTATTGCTAATGATGAAATAATATTAAATTTAGATGCTAATAATAGATAGAACAACAATTC	3152
Qy	616	atctggatgaataacttctcaatcttcaacaccgaaactgtctcaagtccaatatcogaaga	675
Db	3153	ATTGTGATGAAATATTTTCAGTCTGTTTTAATACGGAAATTTAACTCAATCAAAATATTGAAGAA	3212
Qy	676	cggtacaagatccagctctactccgaatacactcgaagactctctgggtgaatcccgctgatg	735
Db	3213	AGATATAAATAATCAATCATATAGCGAATATTTAAAGAGATTTTTGGGGAAATCCTTTAATG	3272
Qy	736	tacacaaagaataactatattgttcaatgctggtgaagaagaactcttatcacaactgaag	795
Db	3273	TACAATAAAGAATATTATATGTTAATCGGGGAATAAAAAATTCATATATTAAACTTAAG	3332
Qy	796	aaagaactctccggtgtgtaaatcctgactcgttccaatacaacacagaactctaaatatc	855
Db	3333	AAAGATTACCTCTAGGTGAAATTTTACACGTACGAATAATATAATCAAAATTTCCAAATAT	3392
Qy	856	atcaactaccgagacctgTacatcggtgaaagttcatcatccgtcgcgaactctaactct	915
Db	3393	ATAAATATAGAGATTTATATATTGAGAAAAATTTATTATAAGAAGAAAGTCAAAATTC	3452
Qy	916	cagttccatcaatgatgacatcgtacgtataaagaagactacatctaccctgacctctccaac	975
Db	3453	CAATCTATAAATGATGATATAGTTTGAAGAAGAGATTTATATATCTAGATTTTTTTTAA	3512
Qy	976	ctgaatcaggaatcgcggtgtatcacacctacaagtactccaagaagaagaagaagctt	1035
Db	3513	TTAATCAAGATGGAGAGTATATATCTATAAATATTTTAAAGAAGAGAGAAAAAATG	3572
Qy	1036	tctctggtccgatctctgtattccgacgaactctcacacaccatcccgatccaagaatac	1095
Db	3573	TTTTTAGCTCCTATAATAGTGAATCTGATGATTTTTTACAATCTATACAAATAAAGAAATAT	3632
Qy	1096	gacgaacacgcgacctactcttctgccagctgtctgttccaagaagaatgaagaatctactgac	1155
Db	3633	GATGAACAGCCACATATAGTTGTCTCAGTTGCTTTTTTAAAAAAGATGAAGAAAGTACGTGAT	3692

Qy	1156	gaatcgcgtcgtgatcgattccacgcgtttctcagaaatcgtgattcgcgaagaatac	1215
Db	3693	GAGATAGGATTGATGGTATTTCATCGTTCTACGAAATCTGGAATTTGATTTAAAGAGTAT	3752
Qy	1216	aaagactactctgcattcccaaatgggtacctgaaggaagttaaacgcgaacccgtacaac	1275
Db	3753	AAAGATTATTTTGTATAGTAAATGGTACTTAAAGAGGTTAAAGGAACCATATATAT	3812
Qy	1276	ctgaaactgggtgtcgaattggcagttcatcccgaaagacgaaggttgagccgaata	1331
Db	3813	TCAAAATTGGGATGTAATTTGGAAGTTTATCTCTAAAGATGAAGGGTGGACTGAATA	3868
RESULT	8		
AF300469			
LOCUS	AF300469	3869 bp DNA linear BCT 08-OCT-2001	
DEFINITION	Clostridium botulinum isolate 13280 type B cryptic neurotoxin-like		
ACCESSION	AF300469	gene, complete sequence.	
VERSION	AF300469.1	GI:15982942	
KEYWORDS			
SOURCE			
ORGANISM	Clostridium botulinum.		
	Clostridium botulinum		
	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
REFERENCE	1 (bases 1 to 3869)		
AUTHORS	Kirma,N., Ferreira,J.L. and Baumstark,B.R.		
TITLE	Characterization of six type A strains of Clostridium botulinum		
JOURNAL	that contain type B toxin gene sequences		
REFERENCE	2 (bases 1 to 3869)		
AUTHORS	Kirma,N., Ferreira,J.L. and Baumstark,B.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-AUG-2000) Department of Biology, Georgia State		
	University, P.O. Box 4010, Atlanta, GA 30302-4010, USA		
COMMENT	NCBI staff are still waiting for submitters to provide appropriate		
	coding region information.		
FEATURES	Location/Qualifiers		
source	1..3869		
	/organism="Clostridium botulinum"		
	/isolate="13280"		
	/db_xref="taxon:1491"		
	/country="USA: Colorado"		
	/note="isolated from peppers in 1972		
	type: A(B)"		
misc_feature	1..3869		
	/note="similar to type B cryptic neurotoxin"		
BASE COUNT	1607 a 365 c 614 g 1283 t		
ORIGIN			
Query Match	51.9%;	Score 696;	DB 1; Length 3869;
Best Local Similarity	71.0%;	Pred. No. 3e-159;	
Matches	935;	Conservative 0;	Mismatches 380; Indels 1; Gaps 1;
Qy	16	aacaatacaatccgaatccgtgacaaatcatctgaacctgcgttcacaaagacaac	75
Db	2554	AATAATATATAGCGATATTTTAAATAATATATCTTAAATTAAGATATAGGGATAAT	2613
Qy	76	aatcgtatcgatctgctcgttgtaacggtgctaaagttgaagtatacgcagcgggttgtaactg	135
Db	2614	AAGTTAAATAGATTTTATCAGGATATGGGGCAAGGTAGAGGTATATGATGGGTCAAGCTT	2673
Qy	136	aatgacaagaaccaagtccaaactgcctcttcgcgttaacttaagatccggtgttactcag	195
Db	2674	AATGATAAAATCAATTTAAATTAAGTACTAGTTCAGCAAAATAGTAAGATTAGTGACTCAA	2733
Qy	196	aatcagaacatacatctccaactccggtattccctggactctctctgtttccctctgattcgt	255
Db	2734	AATCAGATATCATATATTTAATAGTATGTTCCCTTGATTTTAGCGTTAGTGTGTAAGA	2793
Qy	256	atcccgaaatacaagaacgacggtatcccgaaattacatcccaaatgaatacacaccatc	315



Qy	316	aactgatgaagaataaactctggttggaagatctccatccggtgtaaacggtatcatctgg	375
Db	336	AAATGCAATTAATAATCTCGCTGGAAATATATCTATTAGGGGTATAGATAATATG	395
Qy	376	actctgatcatatacaacggtgaagaccacaaactctgtattcttcgaatacaaacatccgtgaa	435
Db	396	ACTTTAACTGATATAAATGGAAAAACCAATCAGTATTTTTTGAATATAGCATAGAGAA	455
Qy	436	gacatctctgaatacatacaatcgtctggttcttcgttacatcacacaataaacctgaacaat	495
Db	456	GATATATCAGACTATATAAATAGATGGTTTTTGTGAATATATACPAATAATTCGGATAAC	515
Qy	496	gctaaatctacatacaacggtgaactgaacttaataccgacatcaaaagacatccgtgaa	555
Db	516	GCTAAATTTATTAATGGTAAGCTAGAACAAATATAGATATTAAGATATAGAGAA	575
Qy	556	gttatcgtaacggtgaatactctcaactctcaactggcgtgacatcgatccaccagttc	615
Db	576	GTTATTCGTAATAGTGAATAATATTTAAATTAGATGCTGATATAGATAGAACACAAATTT	635
Qy	616	atctgatgaataactctccatctcaacacccgaactgtctcagtcacataatcgaa	675
Db	636	ATTGGATGAATAATTTTCAATGATTTTAAACAGAAATTAAGTCAATCAAAATATTAAGAA	695
Qy	676	cgttacaagatccagcttactccgaaacactgaaagactcttggtgtaactccgctgatg	735
Db	696	ATATATAAATCAATCATATAGCAATATTTAAAGATTTTGGGGAAATCCTTTAATG	755
Qy	736	tacaacaagaataactatagtctcaactgctggtgaacagaactcttatcaaacatgaag	795
Db	756	TACAATAAGAATAATATATGTTTAAATCGGGAAATAAAATTCATATATTAACATAAG	815
Qy	796	aaagactccggttggtgaactcgtactcgttccaaatacaacacccgaactctaaataac	855
Db	816	AAAGATTTCATCTGTAGGTGAATTTTACACAGCTAGCAAAATTAATCAAAATTCCAATTAT	875
Qy	856	atcaactccggaactgtacatcgtgtaaggttcacatcgtcgcaaatctcaactct	915
Db	876	ATAAATATAGAAATTTATATATCGGAGAAAAATTTATATAAGAGAAAGTCAAAATCT	935
Qy	916	cagtcacatcaatgatgacatcgtgaagaagactacatctacactcgtactctcaac	975
Db	936	CAATCTATAAATGATGATATAGTTAGAAAAGAAAGATTATATATATCTAGATTTTTTAAT	995
Qy	976	ctgaatcaggaatgctgtatataccatcactcaagtagcttcaagaagaagaagaagctt	1035
Db	996	TCAATCGAGAGTGGAGAGTATATCCCTATAAAGATTTTAAAGGAAGAGAAAAAATG	1055
Qy	1036	ttcctggtccgatctctgattccgacgaactctcaacacccatccgatcaagaataac	1095
Db	1056	GTTTTAGCTAATATATATGATCTTAATGAATTTTACAAACTATACAAATAAAGAAATAT	1115
Qy	1096	gacgaacagccgacactctctgacgctgctgttcaagaagaatgaagaatctactgac	1155
Db	1116	GATGAACAGCCCAACATATAGTTGTCAATTAATCTTTTAAAAAAGATGAAGAAGTCTGAT	1175
Qy	1156	gaactcgtctgatcgttatccacogtttctacgaactcgtgattcgtattcgaagaataac	1215
Db	1176	GAGATAGATTGATGGTATTATCATCGTTTTTACGAATCTGGAATTTGATTTAAAGATTAT	1235
Qy	1216	aaagactctctgatcctccaaatggttacctgaaggaagttaaacgcgaacgctacac	1275
Db	1236	AAAAATTTATTTTGTATAGTAATAGTGGTACTTAAAGAGGTAAGAAAGAACCATATATAC	1295
Qy	1276	ctgaactgggttggaattggcagttcatcc	1306
Db	1296	CCCAATTTAGGTGTAATTTGGCAGTTCATTC	1326
RESULT 10			
CBONTB			
LOCUS	CBONTB	4051 bp	DNA linear BCT 20-APR-1994

DEFINITION C.botulinum type B gene for neurotoxin.  
ACCESSION X71343  
VERSION X71343.1 GI:296148  
KEYWORDS bont/B gene; botulinum neurotoxin type B; neurotoxin type B.  
SOURCE Clostridium botulinum.  
ORGANISM Clostridium botulinum  
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.

REFERENCE 1 (bases 1 to 4051)  
Hutson,R.A., Collins,M.D., East,A.K. and Thompson,D.E.  
Nucleotide sequence of the gene coding for non-proteolytic  
Clostridium botulinum type B neurotoxin: comparison with other  
clostridial neurotoxins  
Curr. Microbiol. 28 (2), 101-110 (1994)  
REFERENCE 2 (bases 1 to 4051)  
Hutson,R.A.  
Direct Submission  
Submitted (06-APR-1993) R.A. Hutson, AFRC Institute of Food  
Research, Reading Laboratory, Microbiology Dept., Earley Gate,  
Whiteknights Road, Reading, RG6 2EF, UK

FEATURES  
source

1. .4051  
/organism="Clostridium botulinum"  
/strain="Exlund 17B ATCC25765"  
/isolate="type B"  
/db\_xref="taxon:1491"  
/clone="PCBOTB1-8"  
83. .91  
97. .3972  
/gene="bont/b"  
97. .3972  
/gene="bont/b"  
/codon\_start=1  
/transl\_table=11  
/product="Bont/B"  
/protein\_id="CAA50482.1"  
/db\_xref="GI:296149"  
/db\_xref="SPTREMBL:Q08077"

1IPRYTGYKPEDFNKSSGIFNRDVCYEDPDYLTNDKKNIFQTLIKLFNRKLSK  
PLGKLEMIINGIPIYGLDRVPLEEFNTIASVTNKLINPGEVERKKGIFANLI  
FGPGVLNENETIDIGQNHFAFEGGIMQKPEYFVSFVNVOVKGASIFNRR  
GYSDPALILMHELIHVLGLYKQVDDLPIVNEKFKFQMSQTDIOAEYLFQGD  
PSIYSPSTKSIYDKVLQNRGIVDRNLKVLVCISDPNININIKYKPKVDEDS  
EGKYSIVSESNKLYKSLMGLFTEINAEYKIKTRASYSDSLPPVKLLDNELI  
TIEGFNRISDKMKYRQNKAKQKAYEISEKHLAVYKQMKCKSVKVGICIDVD  
NENLFFADKNSFSDLSKNRVEYNTQNNYIGNDFPINEILDLTLISKILPSENT  
ESLDFNDVPEYKQPAIKKVTEDNTIFQYLSQTFPLNRIDISLTSFDDALLS  
SKVYSFTSMYIKTKANKWEAGLWAGVQIVDDFVIEANKSSTMDKTDLSLIVPI  
GLALNVGDETAKGNFESAFAEIASGILLFPELLIPVGVFLLESYIDNKNILKTI  
DNALTKRVEKIDMYGLIYAQWLSVTNTQFYTIKEGMYKALNYQAQALEELIKYKNI  
YSEKSNININFINDKNLNGINQAMNDINFINESYSLMKMPLAVKLLDF  
DNTLKLLNLYIDENKLYIGSVEDESKVDKYLKTIIPFSLSTYSNIELIKIPNKY  
NSEILNINIILNRYDNNLIDLSGYKARVEYDVKLNDKQFKLTSSADSKIRVTQ  
QNIIPNSMFLDSVFSVRIPKYNDDIOYTHNEYTIINCMKNSGKISIRGNRII  
WTLIDNGTKTSVFVEYNREDISEYINRWEFTITNLDNAKIVINCTLSNMDIKD  
IGEVIVNGEITPKLDGVDRTQFIMKYFSEFNTQLNQSNIKYIKQSYEYLKDFW  
GNPLMYNKEYTFMAGNSYIKLVKSSVGEILIRSKYNSNINIRNTYIIGERFI  
IRRESQSINDDIVRKEDYIHLDLVHHEMRVYAYKYFKEQEKLFSLISDSNEF  
YKTIKEIKYDEQPSYCOLLEKDEESTDDIIGLIGIHRFYESGVLRRKKYKDYFCISKW  
YLKVKRKPYSNLGCNWFIPKDEGWTE"

BASE COUNT 1679 a 384 c 645 g 1343 t  
ORIGIN

Query Match 49.5%; Score 663.4; DB 1; Length 4051;  
Best Local Similarity 68.5%; Pred No. 2.8e-151;  
Matches 916; Conservative 0; Mismatches 421; Indels 0; Gaps 0;

Qy 3 attcacgatggccaaatacaatccgaaatccctgaaacataatcatcctgaacctgcg 62  
Db 2643 AATAAAATATTATAATAATAATAGCGAAATTTTAAATATATATCTTAAATTAAG 2702



```
Db 2554 ATGTTTAAATATATATACGAAATTTTAAATTAATATATCTTAATTTAAAGATAAAG 2613
QY 70 gacacaaatcgtatcgtctgtgtacggtgctaaagttaaaatgataacgagtggtt 129
Db 2614 GATAATTAATTAATAGATTTATCAGATATGCGGCAAGGTAGAGGTATATGATGAGTC 2673
QY 130 gaactgaatgacagaacacaggttcaaaactgacctctccgctaaactcgaactccggtgtt 189
Db 2674 GAGCTTAAATGATAAAAATCAATTTAAATTAACTAGTTCAGCAAAATAGATGAGTG 2733
QY 190 actcagaatcagaacatcatcttcaactcogtattcctggactctcgtttcctctcg 249
Db 2734 ACTCAAAATCAGAAATATCATATTAATTAATAGTGTCTCTTGATTTTACGCTTACCTTTGG 2793
QY 250 attcgtatcccgaaatcacagaacgagcgttatccagaatcatcacacaaatgaatacacc 309
Db 2794 ATAGAATACCTAAATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2853
QY 310 atcatcaactcgtacgatacgaacgagcgttatccagaatcatcacacaaatgaatacacc 369
Db 2854 ATAGAATACCTAAATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2913
QY 430 cgtgaagacatctcgtgaatacatcatcgtggttctcgtttaccatcaccataaactg 489
Db 2974 AGAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3033
QY 490 aacatgctcaaaatcctacatcaacggttaaaactggaatcctaaactccgacatcaaaagacatc 549
Db 3034 AATAACGCTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3093
QY 550 cgtgaagatcgtacgagtgaaatcaactcgtggttctcgtttaccatcaccataaactg 609
Db 3094 AGAGAAGTATTCTGCTAATGCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3153
QY 610 cagttcatctggatgaaatcctcctcaactcctcaactcctcaactcctcaactcctcaactcct 669
Db 3154 CAATTTATTTGGATGAATATATTTTACGTATTTTAAATGATGATGATGATGATGATGATGATGAT 3213
QY 670 gaagaacggtacagaatccagtcctactcgaataactcgaataactcgaataactcgaataactc 729
Db 3214 GAAGAAAGATATAAAATTTCAATCATATAGCGAATATTTAAAGATTTTGGGGAATTCCT 3273
QY 730 ctgagtacaaagaataactatgttcaatgctggtaacagaaactcttcaatcaaaa 789
Db 3274 TTAATGTACAATAAAGAAATATATATGTTTAAATGCGGGAATAAAAATTCATATATTA 3333
QY 790 ctgaagaagaactcctcgggtgtgtaaatcctgactcgttccaaatcacacacgaactct 849
Db 3334 CTAAAGAAGATTACCTGTAGTGAATTTTAAACAGTAGAGCAAAATATATCAATCAAAATTC 3393
QY 850 aaatacatcaactaccgacacgtgtacatcgttggaagaagtccatcctcgtcgaactct 909
Db 3394 AAATATATAAATATAGAGATTTATATTTGGAGAAAAATTTTATTAAGAAGAAAGTCA 3453
QY 910 aactctcaagtccatcaatgatgacatcgtacgtcgtgaagaagactcacatctacctgga 965
Db 3454 AATTCTCAATCTATAAATGATGATATAGTTAGAAAAAGAGATTATATATATCTAGA 3509
```

## RESULT 12

```
BD009886
LOCUS BD009886 3509 bp DNA linear PAR 31-JAN-2002
DEFINITION Recombinant toxin fragments.
ACCESSION BD009886
VERSION BD009886.1 GI:18638259
KEYWORDS JP 2001502890-A/10.
SOURCE unidentified.
ORGANISM unidentified
```

```
unclassified.
1 (bases 1 to 3509)
Shone,C., Quinn,C.P. and Foster,K.A.
Recombinant toxin fragments
Patent: JP 2001502890-A 10 06-MAR-2001;
MICROBIOLOGICAL RESEARCH AUTHORITY CAMR,THE SPEYWOOD LABORATORY LTD
OS Unidentified
PN JP 2001502890-A/10
PD 06-MAR-2001
PF 22-AUG-1997 JP 1998510524
PR 23-AUG-1996 GB 9617671.4,13-DEC-1996 GB 9625996.5 PI
CLIFFORD CHARLES SHONE,CONRAD PADRAIG QUINN,KEITH ALAN FOSTER PC
CI2N15/31,CI2N1/21,CI2P21/02,C07K14/33,A61K38/16,A61K39/08 CC
Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT CDS 1..3509.
source I..3509
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1468 a 340 c 548 g 1153 t
ORIGIN
```

```
Query Match 39.0%; Score 522.4; DB 6; Length 3509;
Best Local Similarity 71.7%; Pred. No. 7,2e-117;
Matches 685; Conservative 0; Mismatches 271; Indels 0; Gaps 0;
```

```
QY 10 atgcccacaaatcacaaattccgaattcctgaacaataatcatcctgaacctcgttacaacaa 69
Db 2554 ATGTTTAAATATATATAGCGAAATTTTAAATTAATATATCTTAATTTAAAGATAAAG 2613
QY 70 gacacaaatcgtatcgtctgtgtacggtgctaaagttaaaatgataacgagtggtt 129
Db 2614 GATAATTAATTAATAGATTTATCAGGATATGCGGCAAGGTAGAGGTATATGATGAGTC 2673
QY 130 gaactgaatgacagaacacggttcaaaactgactcctccgctaaactcgaactcgtgtt 189
Db 2674 GAGCTTAAATGATAAAAATCAATTTAAATTAACTAGTTCAGCAAAATAGATGAGTG 2733
QY 190 actcagaatcgaacatcatcttcaactcgttatcctcgtgactctcgtttcctctcg 249
Db 2734 ACTCAAAATCAGAAATATCATATTAATTAATAGTGTCTCTTGATTTTACGCTTACCTTTGG 2793
QY 250 attcgtatcccgaaatcacagaacgagcgttatccagaatcatcacacaaatgaatacacc 309
Db 2794 ATAGAATACCTAAATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3153
QY 310 atcatcaactcgtacgatacgaacgagcgttatccagaatcatcacacaaatgaatacacc 369
Db 2854 ATAGAATACCTAAATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2913
QY 370 atcgtgactctgacgatacgaacggttaagacacaaatcgttatcctcgaatacacaactc 429
Db 2914 ATATGACTTTAATGATATATAAATGGAACCAAAATCGGTATTTTGAATAATAACATA 2973
QY 430 cgtgaagacatctcgtgaatacatcatcgtggttctcgtttaccatcaccataaactg 489
Db 2974 AGAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3033
QY 490 aacatgctcaaaatcctacatcaacggttaaaactggaatcctaaactccgacatcaaaagacatc 549
Db 3034 AATAACGCTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3093
QY 550 cgtgaagatcgtacgagtgaaatcaactcgtggttctcgtttaccatcaccataaactg 609
Db 3094 AGAGAAGTATTCTGCTAATGCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3153
QY 610 cagttcatctggatgaaatcctcctcaactcctcaactcctcaactcctcaactcctcaactcct 669
Db 3154 CAATTTATTTGGATGAATATATTTTACGTATTTTAAATGATGATGATGATGATGATGATGATGAT 3213
```



```
RESULT 14
AR000029 AR000029 1330 bp DNA linear PAT 04-DEC-1998
LOCUS Sequence 22 from patent US 5736139.
DEFINITION
ACCESSION AR000029
VERSION AR000029.1 GI:3962560
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1330)
AUTHORS Kink,J.A., Thalley,B.S., Stafford,D.C., Firca,J.R. and Padhye,N.V.
TITLE Treatment of Clostridium difficile induced disease
JOURNAL Patent: US 5736139-A 22 07-APR-1998;
FEATURES
Location/Qualifiers
source
1..1330
/organism="unknown"
BASE COUNT 400 a 339 c 246 g 345 t
ORIGIN

Query Match 19.6%; Score 263; DB 6; Length 1330;
Best Local Similarity 53.9%; Pred. No. 1.5e-53;
Matches 639; Conservative 0; Mismatches 525; Indels 21; Gaps 4;

QY 15 caacaaatacattccgaatactcctgaacaaatacattcctgaacccgtcggttacaaagacaa 74
DB 24 CACTGAATACATCAAGAACATCATCAATACCTCCATCTCGTGAACCTGGCGTACGAAATCCAA 83
QY 75 caactgacgtctgtcgtggtacggtgctaaagttaagatacagacggtgtc----- 129
DB 84 TCACCTGATCGACCTGTCTCGTACGCTCCAAATCAACATCGGTTCTAAAGTTAACTT 143
QY 130 -gaactgaatgacgaagacaggttcaaacgtacctcttcgctaaacttaagatccggt 188
DB 144 CGATCCGATCGAAGAAATCATGATCCAGCTGTTCATCTGGAAATCGTTCACAAATCGAAAT 203
QY 189 tactcagaatcagaacatcatcttcaactccggtattcttcggactctcttcttctctcgt 248

RESULT 15
AR169140 AR169140 1330 bp DNA linear PAT 17-DEC-2001
LOCUS Sequence 22 from patent US 6290960.
DEFINITION
ACCESSION AR169140
VERSION AR169140.1 GI:17906909
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1330)
AUTHORS Kink,J.A., Thalley,B.S. and Stafford,D.C.
TITLE Vaccine and antitoxin for the treatment of C. difficile disease
JOURNAL Patent: US 6290960-A 22 18-SEP-2001;
FEATURES
Location/Qualifiers
source
1..1330
/organism="unknown"
BASE COUNT 400 a 339 c 246 g 345 t
ORIGIN

Query Match 19.6%; Score 263; DB 6; Length 1330;
Best Local Similarity 53.9%; Pred. No. 1.5e-53;
Matches 639; Conservative 0; Mismatches 525; Indels 21; Gaps 4;

QY 15 caacaaatacattccgaatactcctgaacaaatacattcctgaacccgtcggttacaaagacaa 74
DB 24 CACTGAATACATCAAGAACATCATCAATACCTCCATCTCGTGAACCTGGCGTACGAAATCCAA 83
QY 75 caactgacgtctgtcgtggtacggtgctaaagttaagatacagacggtgtc----- 129
DB 84 TCACCTGATCGACCTGTCTCGTACGCTCCAAATCAACATCGGTTCTAAAGTTAACTT 143
QY 130 -gaactgaatgacgaagacaggttcaaacgtacctcttcgctaaacttaagatccggt 188
DB 144 CGATCCGATCGAAGAAATCATGATCCAGCTGTTCATCTGGAAATCGTTCACAAATCGAAAT 203
QY 189 tactcagaatcagaacatcatcttcaactccggtattcttcggactctcttcttctcgt 248
```



Db 204 TATCCTGAAGAAATGCTATCGTATATAAATCTATGTACGAAAACTTCTCCACCTCCTTCTG 263  
Qy 249 gattcgtatcccaagaatacagaacgagcggtatccagaattacataccacaatgaatacac 308  
Db 264 GATCCGTATCCGGAATACCTCAACTCC-----ATCTCTCTGAACAATGAATACAC 314  
Qy 309 catcatcaactgcatgaagaataactctggttggaagatctccatccgcggttaaccggtat 368  
Db 315 CATCATCAACTGCATGGAACAATCTGGTTGGAAGATATCTCTGAACACGCGTGAAAT 374  
Qy 369 catctgactctgatcgatatacaacggttaagaccacaatactctgtattcttcgaatacaacat 428  
Db 375 CATCTGGACTCTGCAGGACTCTCAGGAATCAACACGCGTGTGTATTCAAATACTCTCA 434  
Qy 429 ccgtgaagaacatctctgaatacatcatcgctggttcttccgttaccatccaataa--- 485  
Db 435 GATGATCAACATCTCTGACTACATCAATCGCTGGATCTGTTACCATCACCAACAATCG 494  
Qy 486 cctgaacaatgctaaatctacatacaacggttaaaactggaatcctaataccgacatcaaga 545  
Db 495 TCTGAATTAATCCAAATCTACATCAACGCGCTCTGATCGACCGAAGACCGATCTCAA 554  
546 catccgtgaagtattatcgtaacggtgaataatcttcaaaactggacggtgacatcgatcg 605  
555 TCTGGTAACATCCACGCTTCTAATAACATCATGTTCAAACTGGACGGTGTGCTGACAC 614  
Qy 606 taccagttcatctggtgaataactctccatcttcaacacagaacgtctcagtcacaa 665  
Db 615 TCACCGGTACATCTGGATCAAAATCTTCAATCTGTCGACAAAGAACTGAACGAAAAAGA 674  
Qy 666 tatcgaagaacggtacaagatccagttctactccgaataactcgaagaactcttggttaa 725  
Db - 675 AATCAAGAGCTGTACGACACACAGTCCAAATCTGGGTATCTCTGAAGACATCTGGGGTGA 734  
Qy 726 tccgctgatgtacacaaagaataactatgttcaatgtctggtgtaacagaactcttaccat 785  
Db 735 CTACCTGCAGTACGACAAACCGTACTACATGCTGAATCTGTACGATCCGAACAATAACGT 794  
Qy 786 caaactgaagaagaactctccggttggtgaataactcactgactcgttccaaatacaaccagaa 845  
Db 795 TGACGTCAACAATGTAGGTATCCGCGGTTACATGTACTGAAAGGTCGCGTGGTCTGT 854  
Qy 846 ctctaaatacatacaactaccgacactgtacatcgtgaaagttcatccgctcgcaa 905  
Db 855 TATGACTACCAACATCTACCTGAATCTTCCCTGTACCGTGGTACCAATTCATCATCAA 914  
Qy 906 atctaactctcagttccatcaatgatgacatcgtacgttaaagaagaactacatctacctgga 965  
Db 915 GAAATACGCGTCTGGTAACAAGGACAATATCGTTGCGCAACAATGATCGTGTATACATCAA 974  
966 ctcttcaactgaatcagggaatggcggtatatacacactacaagtacttcaagaagaaga 1025  
975 TGTGTGTATGAAGAACAAAGAAATACCGTCTGGTACCAATGCTTCTCAGGCTGGTGTAGA 1034  
Qy 1026 agaaaagctttctcgtcgtccggtatctcgtgattccgacgaactctacacaccatccagat 1085  
Db 1035 AAAGATCTTGTCTGCTGTGAAATCCCGGAGGTTGGTAAATCTGTCTCAGGTAGTTGTAAT 1094  
Qy 1086 caaagaatacagcaacagccgactactcttgcagctgctgttcaagaagaagaatgaaga 1145  
Db 1095 GAAATCCAAGAACGACGACGGGTA---TCACTAACAAATGCAAAATGAATCTCGAGGACAA 1151  
Qy 1146 atctaactgaagaatcgtgtgatcggtatccaccggtttctacga 1190  
Db 1152 CAATGGTAACGATATCGGTTTCATCGGTTTCCACCGAGTTCAACAA 1196